

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:16:43 : Search time 873 seconds
(without alignments)
8912.548 Million cell updates/sec

Title: US-10-072-621-4
Perfect score: 3455
Sequence: 1 gccaccatgacgagaaaggc.....atgccagtgattgtgtggg 3455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

ched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : NCGeneseq101002:
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2207.8	63.9	2258	21 AAZ94756	Human ATP binding
2	1931.4	55.9	1941	24 AAL44669	Human transporter
3	1062.4	30.7	3376	22 AAK51722	Human polynucleoti
4	1062.4	30.7	3376	22 AAH90113	Human bone marrow
5	1042.2	30.2	3586	24 AB199363	Mouse ischaemic co
6	972.8	28.2	2930	21 AAZ94747	Human ATP binding
7	972.8	28.2	2930	24 ABL63321	Breast cancer rela
8	533.2	15.4	687	24 AB072606	Human MDPT encodin
9	396.8	11.5	2208	23 ABL10101	Drosophila melanog

10	391.8	11.3	2283	23 ABL05663	Drosophila melanog
11	332.8	9.6	2639	23 ABL15499	Drosophila melanog
12	324.2	9.4	2537	21 AAZ98624	Silkworm Bm white
13	313.6	9.1	4497	23 ABL05455	Drosophila melanog
14	253.6	7.3	2182	23 ABL15815	Drosophila melanog
15	243.6	7.1	13815	23 ABL05454	Drosophila melanog
16	238	6.9	287	22 AAK20365	Human brain expres
17	238	6.9	287	22 AAK46460	Human bone marrow
18	238	6.9	287	24 ABS20791	Human genome-deriv
19	220.8	6.4	2396	23 ABL08195	Drosophila melanog
20	220.8	6.4	2590	23 ABL20321	Drosophila melanog
21	220.8	6.4	2894	23 ABL20331	Drosophila melanog
22	220	6.4	2352	23 ABL05135	Drosophila melanog
23	217	6.3	491	22 AAK07701	Human brain expres
24	217	6.3	491	22 AAK33523	Human bone marrow
25	217	6.3	491	24 ABS08376	Human genome-deriv
26	207.6	6.0	2576	23 ABL15497	Drosophila melanog
27	203.2	5.9	5310	23 ABL15498	Drosophila melanog
28	198	5.7	2367	23 ABL22265	Drosophila melanog
29	169.8	4.9	2525	21 AAZ98625	Silkworm Bm white
30	163.4	4.7	2325	23 ABL20329	Drosophila melanog
31	160.6	4.6	8361	23 ABL05662	Drosophila melanog
32	154.6	4.5	1794	23 ABL12777	Drosophila melanog
33	135.6	3.9	5685	23 ABL12776	Drosophila melanog
34	129	3.7	8415	23 ABL15496	Drosophila melanog
35	110.8	3.2	10386	23 ABL22264	Drosophila melanog
36	107.8	3.1	5383	23 ABL20328	Drosophila melanog
37	107.8	3.1	5503	23 ABL20320	Drosophila melanog
38	107.8	3.1	5506	23 ABL08194	Drosophila melanog
39	107.8	3.1	5664	23 ABL20330	Drosophila melanog
40	107.2	3.1	1920	24 ABK51681	DNA encoding human
41	107.2	3.1	2340	24 AAD22009	Human sitosterolae
42	107.2	3.1	2516	24 ABK51682	Human ABCG5 cDNA s
43	105.8	3.1	7869	23 ABL05134	Drosophila melanog
44	105	3.0	11206	21 AAZ47807	Vector for trappin
45	104.2	3.0	8486	24 ABN83687	Vector plasmid pCa

ALIGNMENTS

RESULT 1
AAZ94756
ID AAZ94756 standard: cDNA; 2258 BP.
AC AAZ94756;
XX
XX
DT 01-AUG-2000 (first entry)
XX Human ATP binding cassette cDNA fragment HuWhite2.
XX
XX HuWhite2; ATP binding cassette; human; cholesterol; lipid disorder;
KW atherosclerosis; lipid disorder; dyslipidemia; psoriasis;
KW lupus erythematosus; diagnosis; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200018912-A2.
XX
XX 06-APR-2000.
XX
XX 21-SEP-1999; 99WO-EP06991.
XX
XX 25-SEP-1998; 98US-0101706.
XX (FARB) BAYER AG.
XX Schmitz G, Klucken J;
XX WPI; 2000-293151/25.
XX Adenosine triphosphate binding proteins useful for identifying agents
PT for treating atherosclerosis and other inflammatory disorders .

Db	901	ACCTACCACACCGCGCTGACTTCGTCATCGAGGTGGCCTCTGCGGAGTATGGAGACCTG	960
Qy	967	AACCCCATGTGTTTCAGGGCTGTGCAGATGAGGTGTGCGGTATGGCTTGAGAAGAAGAC	1026
Db	961	AACCCCATGTGTTTCAGGGCTGTGCAGATGAGGTGTGCGGTATGGCTTGAGAAGAAGAC	1020
Qy	1027	AGCCCTGAGAAAGACGAGGTCCCTGCCCCATGCCCTCCTTGTCTCTCGGAAGTGATCC	1086
Db	1021	AGCCCTGAGAAAGACGAGGTCCCTGCCCCATGCCCTCCTTGTCTCTCGGAAGTGATCC	1080
Qy	1087	ATTGAAAGCCACACCTTTGGCCACGACGACCCCTCACACAGTTCTGCATCCTCTTCAAGAGG	1146
Db	1081	ATTGAAAGCCACACCTTTGGCCACGACGACCCCTCACACAGTTCTGCATCCTCTTCAAGAGG	1140
Qy	1147	ACCTTCTGTCCATCCTCAGGACACGGTCTGTACCCACCTACGGTTCATGTGCCACGTG	1206
Db	1141	ACCTTCTGTCCATCCTCAGGACACGGTCTGTACCCACCTACGGTTCATGTGCCACGTG	1200
Qy	1207	GTTATTGGCGTGTCTATCGCGCTCCTCTACTTGCATATTGGCGACGATGCCAGCAAGTTC	1266
Db	1201	GTTATTGGCGTGTCTATCGCGCTCCTCTACTTGCATATTGGCGACGATGCCAGCAAGTTC	1260
Qy	1267	TTCAACAACACCGGCTGCTTCTTCTTCTCATGTCTCATGTTCGCGGCCCTCATG	1326
Db	1261	TTCAACAACACCGGCTGCTTCTTCTTCTCATGTCTCATGTTCGCGGCCCTCATG	1320
Qy	1327	CCAACTGTGCTCACTTCCCTTTAGAGATGGCGTCTTCAATGAGGAGCACTCAACTAC	1386
Db	1321	CCAACTGTGCTCACTTCCCTTTAGAGATGGCGTCTTCAATGAGGAGCACTCAACTAC	1380
Qy	1387	TGTTACAGCCTCAAAAGCGTATTACCTGGCCAAAGACCACTGGCTGACGTGCCCTTTCAGGTG	1446
Db	1381	TGTTACAGCCTCAAAAGCGTATTACCTGGCCAAAGACCACTGGCTGACGTGCCCTTTCAGGTG	1440
Qy	1447	GTGTGTCGGTGTCTACTGACGATGTCTACTGGATGACGGGCCAGCCGCTGAGACC	1506
Db	1441	GTGTGTCGGTGTCTACTGACGATGTCTACTGGATGACGGGCCAGCCGCTGAGACC	1500
Qy	1507	AGCCGCTTCCTGCTCTTCTCAGCCCTGGCCACGCCACCGCCCTTGTGTGGCCCAATCTTTG	1566
Db	1501	AGCCGCTTCCTGCTCTTCTCAGCCCTGGCCACGCCACCGCCCTTGTGTGGCCCAATCTTTG	1560
Qy	1567	GGGCTCGTATCGAGCTGCTTCCAACTCCCTACAGGTGCCCACTTTTGTGGGCCCACTT	1626
Db	1561	GGGCTCGTATCGAGCTGCTTCCAACTCCCTACAGGTGCCCACTTTTGTGGGCCCACTT	1620
Qy	1627	ACCGCCATCCCTGCTCCTTGTCTCCGGCTTCTTGTAGCTTCAAGACCACTCCCACT	1686
Db	1621	ACCGCCATCCCTGCTCCTTGTCTCCGGCTTCTTGTAGCTTCAAGACCACTCCCACT	1680
Qy	1687	TACCTCGAATGGAGTCTCTATCTCCTATGTACAGTATGGCTTTGAGGGTGTGATCCTG	1746
Db	1681	TACCTCGAATGGAGTCTCTATCTCCTATGTACAGTATGGCTTTGAGGGTGTGATCCTG	1740
Qy	1747	ACGATCTATGGATGAGCGAGGAGACCTTGACATGTTTAGAGGAACGCTGCCCGTTC	1806
Db	1741	ACGATCTATGGATGAGCGAGGAGACCTTGACATGTTTAGAGGAACGCTGCCCGTTC	1800
Qy	1807	GAGCCACAGCATCCTCCAGCGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTC	1866
Db	1801	GAGCCACAGCATCCTCCAGCGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTC	1860
Qy	1867	CTGGTCTTGGGATCTTCTTCTAGCCCTGCGCTGCTGCCCTACCTTGTGCTGCTTAC	1926
Db	1861	CTGGTCTTGGGATCTTCTTCTAGCCCTGCGGCTGCTGCCCTACCTTGTGCTGCTTAC	1920
Qy	1927	CGGGTCAAGTCAGAGAGATAG	1947
Db	1921	CGGGTCAAGTCAGAGAGATAG	1941

AAK51722	AAK51722 standard; cDNA; 3376 BP.	
XX	AC	
XX	AC	
XX	AAK51722;	
XX		
DT	06-NOV-2001 (first entry)	
XX		
DE	Human polynucleotide SEQ ID NO 267.	
XX		
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;	
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;	
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;	
KW	nervous system disorder; arthritis; inflammation; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200157190-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	05-FEB-2001; 2001WO-US04098.	
XX		
PR	03-FEB-2000; 2000US-0496914.	
PR	27-APR-2000; 2000US-0560875.	
PR	19-JUN-2000; 2000US-0598075.	
PR	20-JUL-2000; 2000US-0620325.	
PR	01-SEP-2000; 2000US-0654936.	
PR	15-SEP-2000; 2000US-0663561.	
PR	20-OCT-2000; 2000US-0693325.	
PR	30-NOV-2000; 2000US-0728422.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;	
PI	Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;	
PI	Xue AJ, yang Y, Wejhrman T, Goodrich R;	
XX		
DR	WPI: 2001-476283/51.	
DR	P-PSDB: AAM78589.	
XX		
PT	Nucleic acids encoding polypeptides with cytokine-like activities,	
PT	useful in diagnosis and gene therapy -	

	Query Match	30.7%;	Score 1062.4;	DB 22;	Length 3376;
	Best Local Similarity	81.5%;	Pred. No.	1.9e-254;	
	Matches 1389;	Conservative	0;	Mismatches	21; Indels 294; Gaps 2;
Qy	227	CCTGCTGGCGCAAAAGGGTTATAAGACCCTTCTCAAGTCGCCTCTCAGGTAAATTC TGCC	286		
Dd	38	CCCGAGGTCACACAGCTGGTTATAAGACCCTTCTCAAGTCGCCTCTCAGGTAAATTC TGCC	97		
Qy	287	GCGGGAGCTGATTGGCATCATGGGCCCTCAGGGGCTGGCAGCTTGACATTCATGAACA	346		
Dd	98	GCGGGAGCTGATTGGCATCATGGGCCCTCAGGGGCTGGCAGCTTGACATTCATGAACA	157		

347 TCTTGGCAGGATACAGGAGTCTGGAATGAAGGGCAGATCCTGTTAATGAAGGCCAC 406
158 TCTTGGCAGGATACAGGAGTCTGGAATGAAGGGCAGATCCTGTTAATGAAGGCCAC 217
407 GGGAGCTGAGGACCTTCGCGAAGATGCTGCTACATCATGCAAGATGACATGCTGCTGC 466
218 GGGAGCTGAGGACCTTCGCGAAGATGCTGCTACATCATGCAAGATGACATGCTGCTGC 277
467 CCGACCTCACGGTGTGGAAGCATGATGGTCTCTGCTAACTGAATCTTACTGAGAATC 526
278 CCGACCTCACGGTGTGGAAGCATGATGGTCTCTGCTAACTGAATCTTACTGAGAATC 337
527 CCGATGTGAAAACGATCTCGTGACAGATCCTGACGGCATGCGGCCCTGATGCTGCTGCT 586
338 AGGAGGTGAAGAGGAGTGTGACAGAGATCCTGACGGCATGCGGCCCTGATGCTGCTGCT 397
587 CCCACAGGAGACCGCTGCTCTCTGGGGCAGAGGAGGCTCTGGCCATCGCCCTGG 646
398 CCCACAGGAGACCGCTGCTCTCTGGGGCAGAGGAGGCTCTGGCCATCGCCCTGG 457
647 AGCTGGTCAACAAACCGCTGCTGATGTTCTTTGATGAGCCCAACAGTGGTCTGATAGCG 706
458 AGCTGGTCAACAAACCGCTGCTGATGTTCTTTGATGAGCCCAACAGTGGTCTGATAGCG 517
707 CCTCTGTTTCAAGTGTGTCCTCATGAAGTCCCTGGCAGAGGGGGCCGTACCATCA 766
518 CCTCTGTTTCAAGTGTGTCCTCATGAAGTCCCTGGCAGAGGGGGCCGTACCATCA 577
767 TCTGACCATCCACAGCCAGTGTGCAAGTCTTTGAGATGTTTACAGACTCTACATCC 826
578 TCTGACCATCCACAGCCAGTGTGCAAGTCTTTGAGATGTTTACAGACTCTACATCC 625
827 TGAGCCAGGGTCAGTGCATCTTCAAGAGAGTGGTCAACAACTGATCCCTATCAAGG 886
626 -----AGTGATCTTCAAGGGCTGGTCAACAACTGATCCCTATCAAGG 673
887 GACTCGGCTGATGTCGCCACCTTACCAACCCGGCTGACTTCATCATCGAGGTGGCT 946
674 GACTCGGCTGATGTCGCCACCTTACCAACCCGGCTGACTTCATCATCGAGGTGGCT 733
947 CTGGCGAGTATGAGACCTGAAACCCATGTTTTCAGGCTGTGAGATGGGCTGTGG 1006
734 CTGGCGAGTATGAGACCTGAAACCCATGTTTTCAGGCTGTGAGATGGGCTGTGG 793
1007 CTATGGCTGAGAAGAGCAGCCCTGAGAAGACGAGTCCCTGCCCCATGCCCTCTT 1066
794 CTATGGCTGAGAAGAGCAGCCCTGAGAAGACGAGTCCCTGCCCCATGCCCTCTT 853
1067 GTCTCCGGAAGTGGATCCATTGAAGCCACACCTTTGGCCACAGCCCTTCACACAGT 1126
854 GTCTCCGGAAGTGGATCCATTGAAGCCACACCTTTGGCCACAGCCCTTCACACAGT 913
1127 TCTGCATCCTCTTCAAGAGGACCTTCTGCTCATCTCAGGACAGGCTCTGACCCACC 1186
914 TCTGCATCCTCTTCAAGAGGACCTTCTGCTCATCTCAGGACAC - - - - - 959
1187 TACGGTTTATGTCCTCCACGTGGTTATTGGGGTCTCATCGGCTCTCTACCTGCATATTG 1246
960 - - - - - 959
1247 GCGACGATGCCAGAGGCTTCAACACACACCGGCTGCTCTTCTTCCATGCTGTTCC 1306
960 - - - - - 959
1307 TCATGTTCCGCCCTCATGCAACTGCTCACCTTCCCTTAGAGATGGCGGCTTCA 1366
960 - - - - - 959
1367 TGAGGAGCACCTCACTACTGTTACAGCTCAAAGCGTATTACCTGGCCAAAGACCATGG 1426
960 - - - - - 959

QY 1427 CTGACGTGCCCTTTCAGGTGGTGTCTCCGGTGGTCTACTGCAGCATTTGTACTGATGA 1486
Db 960 -----GGTGGTGTCTCCGGTGGTCTACTGCAGCATTTGTACTGATGA 1003
QY 1487 CGGGCCAGCCGCTGAGACCAAGCCGCTTCCCTGCTTCTCAGCCCTGCCACCCACCG 1546
Db 1004 CGGGCCAGCCGCTGAGACCAAGCCGCTTCCCTGCTTCTCAGCCCTGCCACCCACCG 1063
QY 1547 CCTTGGTGGCCCAATCTTTGGGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGGTG 1606
Db 1064 CCTTGGTGGCCCAATCTTTGGGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGGTG 1123
QY 1607 CCACCTTTTGTGGGCCAGTTACCGGCATCCCTGCTTCTTGTTCCTCGGCTTCTTTGTGA 1666
Db 1124 CCACCTTTTGTGGGCCAGTTACCGGCATCCCTGCTTCTTGTTCCTCGGCTTCTTTGTGA 1183
QY 1667 GCTTCAAGACCATCCCACTTACCTGCAATGGAGCTCCTATCTCTCTATGTACAGTATG 1726
Db 1184 GCTTCAAGACCATCCCACTTACCTGCAATGGAGCTCCTATCTCTCTATGTACAGTATG 1243
QY 1727 GCTTGGAGGTGTGATCTCTGACGATCTATGGCATGGAGGAGAGACCTTGACATGTTTATG 1786
Db 1244 GCTTGGAGGTGTGATCTCTGACGATCTATGGCATGGAGGAGAGACCTTGACATGTTTATG 1303
QY 1787 AGGACGCTGCGGCTTCCGGGAGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGATG 1846
Db 1304 AGGACGCTGCGGCTTCCGGGAGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGATG 1363
QY 1847 CCAAGCTCTACATGGACTTCCCTGGTCTTGGGATCTTCTTCTAGCCCTGCGGCTGCTG 1906
Db 1364 CCAAGCTCTACATGGACTTCCCTGGTCTTGGGATCTTCTTCTAGCCCTGCGGCTGCTG 1423
QY 1907 CTAACCTTGTGCTGGCTTACCGGG 1930
Db 1424 CTAACCTTGTGCTGGCTTACCGGG 1447
RESULT 4
AAH90113
ID AAH90113 standard; cDNA; 3376 BP.
XX
AC AAH90113;
XX
DT 01-OCT-2001 (first entry)
XX Human bone marrow cDNA, SEQ ID NO: 470.
DE Human; bone marrow; antineoplastic; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.
XX Homo sapiens.
OS
XX
PN WO200153453-A2.
XX
PD 26-JUL-2001.
XX
PF 23-DEC-2000; 2000WO-US34960.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 30-NOV-2000; 2000US-0250583.
XX (HYSE-) HYSEQ INC.
PA Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI

PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Dmanac RT;
XX WPI: 2001-488707/53.
DR P-PSDB; AAM00994.
XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful
PT for treating e.g. cancer and immune deficiency disorders -
XX Claim 1; Page 592-596; 548pp; English.
PS
CC The present sequence is one of 251 novel human polynucleotides
CC expressed in the bone marrow. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various
CC immune deficiencies and disorders. The deficiencies and disorders may
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous
CC system disorder such as Alzheimer's disease. Detection of the presence
CC or increased expression of the polynucleotide or the protein it
CC encodes is useful for the diagnosis and/or prognosis of one
CC or more types of cancer. The polynucleotide and polypeptide can be
CC used as nutritional sources or supplements and in the screening of
CC chemical compounds as potential drugs.
XX
SQ Sequence 3376 BP; 753 A; 945 C; 896 G; 782 T; 0 other;
Query Match 30.7%; Score 1062.4; DB 22; Length 3376;
Best Local Similarity 81.5%; Pred. No. 1.9e-254;
Matches 1389; Conservative 0; Mismatches 21; Indels 294; Gaps 2;
QY 227 CCTGCTGGCGAAAGGGTTATAGACCCCTTCTCAAGTGCCTCTCAGGTAATTTCTGCC 286
DB 38 CCCGAGGTCACACAGCTGTTATAGACCCCTTCTCAAGTGCCTCTCAGGTAATTTCTGCC 97
QY 287 GCCGGAGCTGATTGGCATCATGGGCCCTCAGGGGCTGGCAAGTCTACATTCATGAACA 346
DB 98 GCCGGAGCTGATTGGCATCATGGGCCCTCAGGGGCTGGCAAGTCTACATTCATGAACA 157
QY 347 TCTTGGCAGGATACAGGAGTCTGAATGAAGGGCAGATCCTGGTTAATGAAGGCCAC 406
DB 158 TCTTGGCAGGATACAGGAGTCTGAATGAAGGGCAGATCCTGGTTAATGAAGGCCAC 217
QY 407 GGGAGCTGAGGACCTTCGCAAGATGTCCTGCTACATCATCAAGATGACATGCTGCTGC 466
DB 218 GGGAGCTGAGGACCTTCGCAAGATGTCCTGCTACATCATCAAGATGACATGCTGCTGC 277
QY 467 CCACCTCACGGTGTGGAAGCCATGATGGTCTCTGCTAACTGAATCTTACTGAGAATC 526
DB 278 CGCAGCTCACGGTGTGGAAGCCATGATGGTCTCTGCTAACTGAATCTTACTGAGAAGC 337
QY 527 CCGATGTGAAAACCATCTGTCGACAGAGATCCTCAGGGCACTGGGCGTGTGCTGCT 586
DB 338 AGGAGGTGAAGAAGAGTGTGTCAGAGATCCTCAGGGCACTGGGCGTGTGCTGCTGCT 397
QY 587 CCCACAGGAGACGCTGCTCTGCGGGCAGAGAGGCTGTCGCCATCGCCCTGG 646
DB 398 CCCACAGGAGACGCTGCTCTGCGGGCAGAGAGGCTGTCGCCATCGCCCTGG 457
QY 647 AGCTGGTCAACAACCCGCTGTCATGTTCTTTGATGAGCCCAAGTGGTCTGATAGCG 706
DB 458 AGCTGGTCAACAACCCGCTGTCATGTTCTTTGATGAGCCCAAGTGGTCTGATAGCG 517
QY 707 CCTCTGTTTCAAGTGTGTCCTCATGAAGTCCCTGCGCAGGGGGCGGTACCATCA 766
DB 518 CCTCTGTTTCAAGTGTGTCCTCATGAAGTCCCTGCGCAGGGGGCGGTACCATCA 577
QY 767 TCTGACCATCCACAGCCCAAGTGTGTCCTTTGAGATGTTTTCACAGCTCTACATCC 826
DB 578 TCTGACCATCCACAGCCCAAGTGTGTCCTTTGAGATGTTTTCACAGCTCTACATCC 625
QY 827 TGAGCCAGGGTCAGTGCATCTTCAAGAGGTGGTCAACCAACCTGATCCCCCTATCTAAAGG 886

DB 626 -----AGTGCATCTTCAAGGGTGGTCAACCACTGATCCCTATCTAAAGG 673
QY 887 GACTCGGCTGCAATGGCCCACTACCAACAACCGGCTGACTTATCATCGAGGTGGCT 946
DB 674 GACTCGGCTGCAATGGCCCACTACCAACAACCGGCTGACTTATCATCGAGGTGGCT 733
QY 947 CTGGGAGTATGAGACCTGACCCCATGTTCTCAGGGCTGTGAGAAATGGCTGTGG 1006
DB 734 CTGGGAGTATGAGACCTGAAACCCCATGTTCTCAGGGCTGTGAGAAATGGCTGTGG 793
QY 1007 CTATGGCTGAGAAGAAGAGCAGCCCTGAGAAGAAGAGTCCCTGCCCATCCCTCTT 1066
DB 794 CTATGGCTGAGAAGAAGAGCAGCCCTGAGAAGAAGAGTCCCTGCCCATCCCTCTT 853
QY 1067 GTCTCCGGAAGTGAATCCATTGAAGCCACACCTTTGCCACAGCACCTCTACACAGT 1126
DB 854 GTCTCCGGAAGTGAATCCATTGAAGCCACACCTTTGCCACAGCACCTCTACACAGT 913
QY 1127 TCTGATCTCTTCAAGAGGACCTTCTCTCCATCTCAGGAGACAGGCTCTGACCCACC 1186
DB 914 TCTGATCTCTTCAAGAGGACCTTCTCTCCATCTCAGGAGACAC----- 959
QY 1187 TACGGTTTCATGTCCACAGTGGTTAATGGCTGTCTATGGCTCTCTACCTGCATATG 1246
DB 960 ----- 959
QY 1247 GCGAGATGCCAGCAAGGTCTTCAACAACACCGGCTGCTCTTCTTCCATGCTTCC 1306
DB 960 ----- 959
QY 1307 TCATGTTCCGCGCCCTCATGCCAAGTGTGCTCAACCTTCCCTTAGAGATGGGGTCTCA 1366
DB 960 ----- 959
QY 1367 TGAGGAGCACCTCAACTACTGGTACAGCCTCAAGCGTATTATCTGCGCAAGACCATGG 1426
DB 960 ----- 959
QY 1427 CTGAGCTGCCCTTTCAGTGGTGTGTCGCGGTGCTACTGACAGATTTGTACTGGATGA 1486
DB 960 -----GGTGTGTGTCGCGGTGCTACTGACAGATTTGTACTGGATGA 1003
QY 1487 CGGGCAGCCGCTCAGACACAGCCGCTTCTCTCTCAGCCCTGCCACCGCACCG 1546
DB 1004 CGGGCAGCCGCTCAGACACAGCCGCTTCTCTCTCTCAGCCCTGCCACCGCACCG 1063
QY 1547 CTTGGTGGCCCAATCTTTGGGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGTGG 1606
DB 1064 CTTGGTGGCCCAATCTTTGGGGCTGCTGATCGGAGCTGCTTCCAACTCCCTACAGTGG 1123
QY 1607 CCACCTTTTGGGGCCCAAGTTACCGCCATCCCTGCTCTGTTCTCCGGCTCTTTGTCA 1666
DB 1124 CCACCTTTTGGGGCCCAAGTTACCGCCATCCCTGCTCTGTTCTCCGGCTCTTTGTCA 1183
QY 1667 GCTTCAAGACATCCCACTTACCTGCAATGGAGTCTCTATCTCTCTATGTACAGTATG 1726
DB 1184 GCTTCAAGACATCCCACTTACCTGCAATGGAGTCTCTATCTCTCTATGTACAGTATG 1243
QY 1727 GCTTTGAGGGTGTGATCTCTGAGATCTATGGCATGGAGGAGAGACCTGACATGTTTAG 1786
DB 1244 GCTTTGAGGGTGTGATCTCTGAGATCTATGGCATGGAGGAGAGACCTGACATGTTTAG 1303
QY 1787 AGGAACGCTGCCGTTCCGGAGCCACAGAGCATCCTCCGAGCGCTGGATGTGGAGATG 1846
DB 1304 AGGAACGCTGCCGTTCCGGAGCCACAGAGCATCCTCCGAGCGCTGGATGTGGAGATG 1363
QY 1847 CCAAGCTTACATGACTTCTCTGGTCTTGGGATCTCTTCTAGCCCTCGGGCTGCTGG 1906
DB 1364 CCAAGCTTACATGACTTCTCTGGTCTTGGGATCTCTTCTAGCCCTCGGGCTGCTGG 1423
QY 1907 CCTACCTTGTGCTGGGTTTACCGGG 1930
DB 1907 CCTACCTTGTGCTGGGTTTACCGGG 1930

Db	1460	ACTGTTCTGACCTTCCCTGGAGATGAGTGTCTTCTCCGGAGACCTGCAACTACTGG	1519
Qy	1390	TACAGCCTCAAGCGTATTACCTTGGCCAAAGACCATGGCTGACCTGCCCTTTACAGTGGTG	1449
Db	1520	TACAGCCTGAAGCGCTACTACCTGGCAAAGACCATGCCGATGTCCTCTTTCAGATCATG	1579
Qy	1450	TGTCGCGGTGCTACTTCGACGACATTGTACTGATGACGGGCCAGCCCGCTGAGACCAGC	1509
Db	1580	TTCCCTGTGGCCTACTGCGAGTATCGTATATCTGATGACGTCACGACCGCTCGGACGCTGTG	1639
Qy	1510	CGCTTCCTGCTCTTCTCAGCCCTGGCCACCGCCACCGCCTTGGTGGGCCCAATCTTTGGGG	1569
Db	1640	CGTTTTGTGCTGTGCTGCTGCTGGTACCATGACATCGCTGCTGGCCCAAGTCTTAGGA	1699
Qy	1570	CTGCTGATCGGAGCTGCTTCCAACTCCCTACAGGTGGCCACTTTTGTGGGCCCAGTTACC	1629
Db	1700	CTACTGATTGGAGCTGCATCCACATCCCTGCAGGTTGCGACATTTGTGGTCCCGTGACA	1759
Qy	1630	GCCATCCCTGTCCTCTTGTCTCCGGCTCTTTTGTGAGCTTCAAGACCATCCCACTTAC	1689
Db	1760	GCCATCCCGTCTGCTCTCTCCGGATTCTTTGTGAGCTTTGACACCATCCAGCCTAC	1819
Qy	1690	CTGCAATGAGACTCCTCATCTCTCCTATGTACAGTATGGCTTGGAGGTGATCCTGACG	1749
Db	1820	CTGCATGTGATGCTTACATCTCCTATGTACAGATCGGCTTTGAGGGGGTCACTCTGTCC	1879
Qy	1750	ATCTATGGGATGGAGCGAGGAGACCTTGACATG--TTTAGAGGAACGCTGCCCGTTCCCG	1806
Db	1880	ATCTACGGCTTGAGCCGAGAGACCTGCATCGGACATCGGGAGACATGCCACTTCCAG	1939
Qy	1807	GAGCCACAGACATCCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTC	1866
Db	1940	AAGTCAAGGCCCATCCTGAGGAGCTGGAGCTGGAAATGCGAAGCTGTAACCTGGATTTT	1999
Qy	1867	CTGGCTTGGGCATCTTCTTCTTCTAGCCCTGCGGCTGCTGGCCTTACCTTGGCTGCCTTAC	1926
Db	2000	ATCGTCTGGGCATCTTCTTCATCTCCCTGGGCTCATCGCCTATTTTGTCTTCACATAC	2059
Qy	1927	CGGGTCAAGTCAGAGAGATAGAGGCTTGGCCCCAGCCTGTACCCAGCCCTCGAC	1981
Db	2060	AAATCCGGGCTGAGAGGTAAAAACCCGCTTCAAGCCAGCAACGAGGCAAGAC	2114

RESULT 6
AAZ94747
ID AAZ94747 standard; cDNA; 2930 BP.
XX
XX AC
XX AAZ94747;
01-AUG-2000 (first entry)
Human ATP binding cassette ABCG1 (ABCG8) cDNA.
XX
XX ABCG1; ABC8; ATP binding cassette; human; phagocyte; cholesterol;
XX phospholipid; transporter; inflammation; atherosclerosis;
XX lipid disorder; dyslipidemia; psoriasis; lupus erythematosus;
XX diagnosis; gene therapy; chromosome 21q22.3; ss.
XX
XX OS
XX Homo sapiens.
XX
XX PN
XX WO200018912-A2.
XX
XX PD
XX 06-APR-2000.
XX
XX PF
XX 21-SEP-1999; 99WO-EP06991.
XX
XX PR
XX 25-SEP-1998; 98US-0101706.
XX
XX PA
XX (FARB) BAYER AG.
XX
XX PI
XX Schmitz G, Klucken J;
XX

WPI: 2000-293151/25.

Adenosine triphosphate binding proteins useful for identifying agents for treating atherosclerosis and other inflammatory disorders -

Claim 9; Page 129-130; 154pp; English.

The present sequence is that of human cDNA encoding ATP binding cassette protein ABCG1 (ABCG8). The cDNA was identified using a differential display method in which monocytes from peripheral blood were subjected to macrophage differentiation and cholesterol loading with acetylated low density lipoproteins and subsequent deloading with high density lipoprotein (HDL3) to identify cholesterol sensitive genes. The ABCG1 gene maps to human chromosome 21q22.3. ABCG1 was identified as a cholesterol sensitive switch. It is up-regulated by macrophage colony stimulating factor dependent phagocytic differentiation, and expression is massively induced by cholesterol loading and almost completely set back to differentiation dependent levels by HDL3. ABCG1 is the human homologue of the *Drosophila* white gene. Sequencing of the promoter (see AA2947784) revealed important transcription factor binding sites relevant for phagocytic differentiation and lipid sensitivity. ABCG1 was also identified as a cholesterol transporter and modulator of choline-containing phospholipids (phosphatidylcholine, sphingomyelin). The invention provides cholesterol-sensitive ABC genes (see AA2947734-63) that can be used for diagnostic and therapeutic applications, and for biochemical or cell-based assays to screen for pharmacologically active compounds useful for the treatment of lipid disorders, atherosclerosis or other inflammatory diseases such as psoriasis and lupus erythematosus.

Sequence 2930 BP; 665 A; 770 C; 780 G; 715 T; 0 other;

Query Match	28.2%	Score 972.8	DB 21:	Length 2930
Best Local Similarity	70.5%	Pred. No. 4e-232		
Matches 1351: Conservative	0	Mismatches 517	Indels 48	Gaps 2
QY	81	GGAGGGGGCGGAACCCCTGTGCTGACCAACGCGCACCTGGAAGAGGTGGAGAACACACATCAC	140	
DB	141	GGAGGCCATGACAGCGAGCTCTGTAATGGACATCTGAAAAAGTAGATAAATACCTCAC	200	
QY	141	TGAAGCCCCAGCGCTTCTCCACCTGCCCAAGCCCTCAGCCGTGGAGCATCGAGTTCGTGGA	200	
DB	201	GGAAGCCCAAGCGCTTCTCTCTCTGCTCGGAGGGCAGCTGTGAACATTGAATTCAGGA	260	
QY	201	GCTGTCTATTTCGTGCGGGAGGGCCCTGCTGGCGCAAAAGGGGTATTAAGACCCTTCT	260	
DB	261	CCTTTCTATTGGTTTCCTGAGGACCTGGTGGAGGAAGAAGGATACAAGACCCTCCT	320	
QY	261	CAAGTCCCTCTCAGGTAAATCTGCCGCCGGGAGCTGATTGSCATCATGGGCCCTTCAGG	320	
DB	321	GAAAGGAATTCGGGGAAGTCAATAGTGGTAGTTGGTGGCCATTATGGGTCTTCCCG	380	
QY	321	GGCTGCAAGTCTACATTTCATGAACATCTTTGGCAGGATACAGGGAGCTCTGGAATGAAGG	380	
DB	381	GGCCGGGAGTCCACGCTGATGAACATCTTGGCTGATACAGGAGACGGGCATGAAGG	440	
QY	381	GCAGATCTCTGGTTAATGGAAGGCCACGGAGCTGAGGACCTTCCGCAAGATGCTCTGCTA	440	
DB	441	GGCCGTTCCTCATACAGGGCTGCCCGGACCTGCGCTGCTTCCCGAAGGTGCTCTGCTA	500	
QY	441	CATCATCGAAGATGACATCTGCTGCCGCACTCAGGTGTTGGAGACCATGATGCTCTC	500	
DB	501	CATCATGCAAGATGACATCTGCTGCCGATCTCATCTGTGCAAGGAGGCCATGATGCTGTC	560	
QY	501	TGCTAACCTTGAATCTTCTACTGAGAAATCCCAGATGTGAAAAACGATCTCGTGACAGAGATCCT	560	
DB	561	GGCACAATCTGAGCTTCAGSAGAGGATGCAAGCAGAGGGAATGCTCAAGGAGATACT	620	
QY	561	GACGGCACTGGGCCATGATCTGTGTCCTCCACAGGAGCAGACCCCTGCTCTCTGGCGGCA	620	
DB	621	GACAGCGCTGGGCTTGTCTCTTGTGCCAACACCGCGAGCGGGAGCTGTGACGTGCTCA	680	

Qy	621	GAGGAAGCGTCTGGCCATCGCCCTGGAGCTGGTCAACAACCCGCTTGCTATGTTCTTTGA	680
Db	681	GCSCAAGCGCTGGCCATCGCGCTGGAGCTGGTGAACAACCCCTCCAGTCATGTTCTTGA	740
Qy	681	TGAGCCACCAAGTGTCTGGATAGCGGCTTGTGTTCCRAAGTGGTCTCCCTCATGAATC	740
Db	741	TGAGCCACCAAGCGGCTGGACAGCGGCTCTGTTCCAGGTGGTCTCGCTGTAAGAAAG	800
Qy	741	CTTGGCACAGGGGGCGGTACCATCATCTGCACCATCCACACCCAGCTGCCAAGCTCTT	800
Db	801	GCTCGCTCAAGGGGTGCTCCATCATTTGCCACCATCCACAGCCCGCCAACTCTT	860
Qy	801	TGAGATGTTTGACAAGCTCTACATCTGAGCCAGGTCAAGTGCATCTTCAAGAGAGTGT	860
Db	861	CGAGCTGTTTCGACAGCTTACGCTCTGAGTCAAGGACAAATGTGTACCGGGGAAAAGT	920
Qy	861	CACCAACCTGATCCCTCATCTAAAGGAGCTCGGCTTGCATTTGCCCCACCTACCAACCC	920
Db	921	CTCAATCTTGTGCCATATTTGAGGATTTGGGTCTGAACGTGCCAACCTTACCACAACCC	980
Qy	921	GGCTGATCTTCATCATCGAGGTGGGCTCTGGCAGTATGGAGACTGAACCCCATGTTGT	980
Db	981	AGCAGATTTGTTCATGGAGGTTGCATCCGCGAGTACGGTGTATCAGAACAGTCCGCTGT	1040
Qy	981	CAGGGCTGTGCAAAATGGGCTGTGCCTATGCTGAGAAGAGCAGCCCTCAGAAAGAA	1040
Db	1041	GAGAGCGGTTCCGGAGGCGATGTGTACTCAGACCAACAAGAGAGACTCGGGGGTGATGC	1100
Qy	1041	CGAGGTCCTTGGCCCCATGCCCTCCTTGTCTCTCGGAAGTGGATCCCATTTGAAA-----	1093
Db	1101	CGAGGTGAACCCCTTTCTTTTGGCACCGCCCTCTGAAGAGTTAAAGCAGACAACAGATT	1160
Qy	1094	-----GCACACCTTTTCCACACGACAC	1115
Db	1161	AAAGGGTTTCAAGAAAGACTCTCTGTCATGGAAGGCTGCCACAGCTTCTCTCCAGCTG	1220
Qy	1116	CCTCACACAGTTCTGCATCCTTCAAGAGGACTTCTCTGTCATCCTCAGGACACGGT	1175
Db	1221	CCTCAGCAGTTCTGCATCCTTCAAGAGGACTTCTCTCAGCATCATGAGGACCTGGT	1280
Qy	1176	CCTGACCCACCTACGTTTCATGTCCACAGCTGTTATTTGGGTGCTCATCGGCTCTCTCTA	1235
Db	1281	CCTGACACACTCGCATCACTTCGACATTTGGGATCGGCTCTCTCATTTGGCTCTGTGA	1340
Qy	1236	CCTGCATATTTGGCGAGATGCCAGCAGGTCTTCAACAACACCGGCTGCTCTCTTCTCTC	1295
Db	1341	CTTTGGGATCGGAACGAACCAAGAGGTCTTTGAGCACTCCGGCTTCTCTTCTCTC	1400
Qy	1296	CATGCTGTTTCTCATGTTCCGCGCTCATGCAACTGTCTCACCTTCCCTTACAGAT	1355
Db	1401	CATGCTGTTTCTCATGTTCCGCGCTCATGCTACTGTTCTGACATTTCCCTCGGAGAT	1460
Qy	1356	GGCGGTCTTCATGAGGAGACCTCAACTACTGGTACAGCCCTCAAGCGTATTACCTGGC	1415
Db	1461	GGAGTCTTTTCTTTCGGGAACACCTGAACTACTGGTACAGCTCAAGGCTTACTACCTGGC	1520
Qy	1416	CAAGACCATGGCTGAGTGGCTTTCAGGTGGTGTGTCGGGTGGTCTACTGCAGCATGT	1475
Db	1521	CAAGACCATGGCAGAGTGGCTTTCAGATCATGTTCCCAAGTGGCTTACTGCAGCATCGT	1580
Qy	1476	GTACTGATGACGGGCGCCGCTGAGACAGCGGCTTCTGCTCTTCTTCAGCCCTGGC	1535
Db	1581	GTACTGATGACGTGCGACCGTCCGACGCGGTGCTGTTGTGCTGTTCCGCGCTGG	1640
Qy	1536	CACGGCACCGCTTGTGTGGCCAACTTTTGGGCTGCTGATCGAGCTGCTTCCAACTC	1595
Db	1641	CACCATGACCTCCTGTTGGCACAGTCCCTGGGCTTCTGATCGGACCGCTCCACGTC	1700
Qy	1596	CCTACAGGTGGCACATTTGTGGGCCAGTTACGGCCATCCCTGTCTCTTGTCTCCGG	1655
Db	1701	CCTGCAAGTGGCACATTTCTGTGGGCCAGTGACAGCCATCCCGTGTCTCTGTTCTCGG	1760

PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 02-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 03-OCT-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244867P.
 PR 01-NOV-2000; 2000US-245084P.
 XX (AVAL-) AVALON PHARM.
 XX
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX
 DR WPI; 2002-188264/24.
 DR
 XX Screening for anti-neoplastic agent involves exposing cells to a
 PS chemical agent to be tested for anti-neoplastic activity, and
 XX determining a change in expression of a gene of a signature gene set
 PS
 XX Claim 1; SEQ ID 1658; 44pp; English.
 XX
 XX The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical
 CC structure and/or properties of the agent. M1 can be used in the
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilms' tumour.
 XX
 XX Sequence 2930 BP; 665 A; 770 C; 780 G; 715 T; 0 other;
 SQ

Query Match 28.2%; Score 972.8; DB 24; Length 2930;

Best Local Similarity .70.5%; Pred. No. 4e-232;

Matches 1351; Conservative 0; Mismatches 517; Indels 48; Gaps 2;

81 GGACGGGGGGAACCCCTGTGCTGACACGACCTGAAGAAGGTGGAGAACCCACATCAC 140
 141 GGAGGCCATGAGAGGACCTGCTGAATGGACATCTGAAAAGTAGATAATACCTCAC 200
 141 TGAAGCCCAAGCCTTCTCCACCTGCGCAAGCGCTCAGCGGTGACATCGAGTTCTGGA 200
 201 GGAAGCCCAAGCCTTCTCCCTGCTGCGGAGGCGCTGTGAACATTGAATTGAGGA 260
 201 GCTGTCTATTCCGTGCGGGAGGGCCCTGCTGCGGCAAAAGGGGTATTAAGACCCCTT 260
 261 CCTTTCTTATTCCGTGCTGGAAGGACCTTGTGTGGAGGAAGAAGATACAAGACCCCT 320
 261 CAAGTGCCTCTCAGGTAAATTCGCGCGGGAGCTGATGGCATCATCGGGCCCTCAGG 320
 321 GAAAGAAATTCGGGAAGTTCAATAGTGGTGGTGGGCAATATGGGTCCTTCGGC 380
 321 GCGTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGGAGTCTGGAATGAAGG 380
 381 GCGCGGAAGTCCACGCTGATGAACATCTGCTGATACAGGAGCGGATGAGG 440
 381 GCAGATCTCTGTTTATGGAAGCCACGGGAGCTGAGGACCTTCCCGAAGATGCTCTGTA 440

441 GGCCGCTCTCATCAACGGGCTGCCCGGACCTGCGCTGCTCCGGAAGGTGCTCTGCTA 500
 441 CATCATGCAAGATGACATGCTGCTGCCGACACCTCAACGGTGTGGAAGCCATGATGCTCTC 500
 501 CATCATGCAAGATGACATGCTGCTGCCGATCTCACTGTGAGGAGGCGCATGATGCTGTC 560
 501 TGTAACTCTGAATCTTACTGAGAATCCCGATGTGAAAAACGATCTCCTGACAGAGATCCT 560
 561 GCCACATCTGAAGCTTCAGGAGAGGATGAAGCAGAGGAAATGCTCANGAGATACT 620
 561 GAGGCACTGGGCGCTGATGCTGCTCCACAGAGGACAGCCCTGCTCTCTGGCGGCA 620
 621 GACAGCCTTGGGCTTGTCTGCTGCGCAACACGCGGAGCGGAGCTGTCAAGTCTGCTCA 680
 621 GAGGAGCGTCTGGCCATCGCCCTGAGCTGGTCAACAACCCGCTGCTGCTCTTTTCA 680
 681 GGGCAAGCGCTGGCCATCGCGTGGAGCTGGTGAACAACCCCTCCAGTCTGCTTCTCGA 740
 681 TGAGCCCAACAGTGGTCTGATAGCGCTCTTCTTCAAGTGGTGTCTCCTCATGAAGTC 740
 741 TGAGCCCAACAGCGGCTGACAGCGCTCTGCTTCCAGGTGGTCTCCTGATGAAGG 800
 741 CTGGCAGAGGGGGCGCTTACCATCATCTGCACCATCCACAGCCAGTGCACCAAGCTCT 800
 801 GCTCGCTCAAGGGGTGCTCTCCATCATTTGCACCATCCACAGCCAGCCGCAAACTCT 860
 801 TGAGATGTTTGACAGCTCTACATCTGAGCCAGGCTCAGTGCATCTTCAAGGAGTGT 860
 861 CGAGCTGTTTGCAGCAGCTTTACGCTGTAGTCAAGGACAAATGTGTACCGGGGAAAGT 920
 861 CACCAACCTGATCCCTATCTAAAGGACCTCGGCTTGCATTTGCCCCACCTACACAAACC 920
 921 CTGCAATCTGTGCCATATTTGAGGATTTGGTCTGAACCTGCCCAACCTACCAACACC 980
 921 GCGTACTTATCATCGAGTGGGCTGCGGAGTATGAGACCTGGAACCCCATGTTGT 980
 981 AGCAGATTTGTGATGGAGTTCATCGCGGAGTACGCTGATCAGAACAGTCCGCTGCT 1040
 981 CAGGCTGTGCAAGTGGCTGTGGCTATGCTGAGAGAAGAGAGAGAGCTCGGGGGTATGC 1040
 1041 GAGAGCGTTCGGGAGGATGTGTGACTCAGACCAAGAAGAGAGAGCTCGGGGGTATGC 1100
 1041 CGAGTCCCTGCGCCCATGCGCTTCTTGTCTCCGGAAGTGGATGCCATTTGAAA----- 1093
 1101 CGAGGTGAACCTTTTCTTTGGCACCGCCCTCTGAAGAGGTAAAGCAGACAAACGATT 1160
 1094 -----GCCACACCTTTGCCACACAGC 1115
 1161 AAAGGGTTTGAGAAAGGACTCTCGTCCATGGAAGGCTGCCACAGCTTCTCTGCCAGCTG 1220
 1116 CCTCACAGTTCGATCTCTTCAAGAGGACCTTCTCTGTCATCTCAGGAGACAGCT 1175
 1221 CCTCAGCAGTTCGATCTCTTCAAGAGGACCTTCTCAGCATCATGAGGAGCTCGGT 1280
 1176 CCTGACCCACCTACGTTTCATGTCACCGTGGTATTGCGGTGCTCATCGGCTCTCTCTA 1235
 1281 CCTGACACACCTGCGCATCACCTCGACATTTGGATCGGCTCTCTCATTTGGCTGCTGA 1340
 1236 CTGATCATTTGGCGAGATGCGAGCAAGTCTTCAACAACACCGGCTGCTCTCTCTCTC 1295
 1341 CTGTTGGGATCGGGAACGAAACCAAGAAGTCTTTAGCAACTCTCGGCTCTCTCTCTCTC 1400
 1296 CATGCTGTCTCATGTTTCCGCGCCCTCATGCAACTGTGCTCACTTCCCTTACAGAT 1355
 1401 CATGCTGTCTCATGTTTCCGCGCCCTCATGCTCTACTTGTGTGACATTTCCCTGAGAT 1460
 1356 GCGGTCTTTCATGAGGAGGACCTCAACTACTGTGTGACAGCTTCAAGAGCTTATTAACCTGC 1415
 1461 GGGAGTCTTCTTCGGGAACACCTGAACACTGTTGTTACAGCTTGAAGGCTTACTACCTGC 1520
 1416 CAAGACCATGGCTGACGTCCTTTTCAAGTGTGTGTCGGGTGCTTACTGACAGATGT 1475
 1521 CAAGACCATGGCAGACGTCCTTTTCAAGTGTGTGTCGGGTGCTTACTGACAGATGT 1580

Db	1158	GCCTCTGCGATGTACCAACAATCCAGCGGATTACATTATCGAGCTCGCTTGGGAGAGTA	1217
Qy	957	TGGAGACCTGAACCCCATGTTCTTTCAGGCGCTGTCAGAAATGGCGTGTGCGTATGCGTGA	1016
Db	1218	TGGATACGACAAGGTCGACACCTTAAAGCTCGCCACTGAGAACGAACTGCCTGACCTTG	1277
Qy	1017	GAAGAAGACAGCCCTGAGAAGACGAGTCCCTGCCCATGCCCTCTTGTCTCTCCGGA	1076
Db	1278	GTTCATAATCCAGTGCAGGTGTGGCGCTGAGGTCTTGATGAGAAAGTATCCCATACC	1337
Qy	1077	AGTGGATCCATTGAAGACCAACACTTTTGCCACAGCACCCCTCACACACTTCTTGCATCCT	1136
Db	1338	AAAGAAACCAAAAGCCGATCCCTGGAGGACACAGCTACTCAAACCACTGCTCTGTGCT	1397
Qy	1137	CTTCAAGAGACCTTTCCTGCTCCATCCTCAGGAGACAGGTCTCTGACCCACCTACAGGTTCA	1196
Db	1398	TTTGGCAGCAGGTTTCATCAAAAGCCAAACGAGACGACACCACTGACGCACTTGAGGATCGG	1457
Db	1197	GTCCCACTGGTTATTGGCGTCTCATCGGCTCCTCTACCTGTCATATTGGGACGATGC	1256
Db	1458	AGTCACATAGCAGTGGCGCTCTGTTCCGAGCTATGTACGATCACGGGACGTGAAGG	1517
Qy	1257	CAGCAAGCTTTCACACACACCGGTGCTCTTCTTCCATGCTGTTCCTCATCTTCCG	1316
Db	1518	ATCACGAGTGTGACAACTACAATCTTCTGTTCCTACTGATGCACCATTTCCATGAC	1577
Qy	1317	CGCCCTCATGCCAAGCTGCTCACCTTCCCTTAGAGATGGCGGCTTTCATGAGGGAGCA	1376
Db	1578	CAGATGATGTTGACATTTCTGACTTTCCTCCCATGGACATCTCCATATTGATCAAGGAGCA	1637
Qy	1377	CCTCAACTACTGGTACAGCCTCAAAGCGTATTACCTGGCCAAAGACCATGGCTGAGTGCC	1436
Db	1638	CTTCAATCGCTGCTACTACTCAAGCCTACTACACTGCCATGACACTCTGGACCTGCC	1697
Qy	1437	CTTTCAGGTGGTGTGTCGGTGGTCTACTGACAGCATTTGTGTACTGGATGACGGGCCAGCC	1496
Db	1698	CATATCTATCATAAAGTTGTTTCTCTTCACTGTCATCGCTTTACCTTTGGAGTTATCAGCC	1757
Qy	1497	CGCTGAGACAGCGGCTTCCCTCTCTTCACGCGCTGGCCACCGCCAGCTTGGTGGC	1556
Db	1758	GATGGAGTGATCAGATTCTTATGTGTTCTTCCATTTAGTTTACTGACCGCTCTTTGGGG	1817
Qy	1557	CCAATCTTTGGGCTGCTGATCGAGCTGCTTCCAACTCCCTACAGGTGGCCACTTTGT	1616
Db	1818	ACATAGTTTGGTTTGATGATTGG---TGCCCTGGTTGCGATGTGGTGAATGGAACGCTCT	1874
Db	1617	GGCCCCAGTTACCGCCATCCCTGTCTCTTCTTCGCGCTCTTTGTACAGCTTCAAGAC	1676
Db	1875	GGCCCCAGTGCTAACGATTCCCATGATGATGTTCGCGCGCTTTGGAGTGACCTCCGCTGA	1934
Qy	1677	CATCCCCACTTACTGCAATGGAGTCCCTATCTCTCTATGTCAGGTATGGCTTTGAGGG	1736
Db	1935	TCTGCCAAGCTACTTAAAGCTGGGCGAGTCACATATCATACTTAAAGATATGGACTTGAGGG	1994
Qy	1737	TGTGATCTGACGATCTATGGCATGGAGCGAGAGACCTTGACATGTTTAGAGAAC---G	1793
Db	1995	CTTCTATCAGCCATTATGGCTTGGATCGAGTACCTTGGCTCGCAGGAGCGGCCGTA	2054
Qy	1794	CTCCCGTTCCGGGACCAAGACATCTCCGAGCGCTGGATGTCGAGGATGCCAAGCT	1853
Db	2055	CTGCCATTACAGGTATCCAAAGAAATCTTATAGAGAAATCACCATGAGGGGTGATCAGTT	2114
Qy	1854	CTACATGGACTTCTCGTCTTGGGCTCTTCTTCTTAGCCCTGCGGCTGCTGGCCCTACT	1913
Db	2115	CTGGAACGATGTATCGCTGGGAGTCATGATCTCGTGTTCCTCGATTTGTGCTCTACGT	2174
Qy	1914	TGTGCTCGTTACCGGGTCAAGTCAGAGAGATAG	1947
Db	2175	GGTGTCTGAGGGCCAAAGTCAAGTCGATCCGATAG	2208

ID	ABL05663 standard; cDNA; 2283 BP.
XX	
XX	ABL05663;
XX	
XX	26-MAR-2002 (first entry)
DT	
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 11471.
XX	
XX	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
KW	
XX	Drosophila melanogaster.
OS	
XX	WO200171042-A2.
PN	
XX	27-SEP-2001.
PD	
XX	
XX	23-MAR-2001; 2001WO-US09231.
PF	
XX	23-MAR-2000; 2000US-191637P.
PR	
PR	11-JUL-2000; 2000US-0614150.
XX	
XX	(PEKE) PE CORP NY.
PA	
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
DR	P-PSDB; ABB61560.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
XX	Claim 1; SEQ ID NO 11471; 2lpp + Sequence Listing; English.
PS	
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences .

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:37:55 ; Search time 8632 Seconds
(without alignments)
11648.540 Million cell updates/sec

Title: US-10-072-621-4
Perfect score: 3455
Sequence: 1 gccaccatggcggaagc.....atgcagttgtattgtggg 3455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vt.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vt.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3426.6	99.2	3498	9	HSA308237	AJ308237 Homo sapi
2	3424.6	99.1	3501	9	HS300465	AJ300465 Homo sapi
3	2761	79.9	2838	9	HS802306	AL137563 Homo sapi
4	2133.4	61.7	3474	10	AF411084	AF411084 Mus muscu
5	2132.2	61.7	3522	10	BC016200	BC016200 Mus muscu
6	2125	61.5	3520	10	MMU426047	AJ426047 Mus muscu
7	1741.4	50.4	157405	2	AP001315	AP001315 Homo sapi
8	1732.8	50.2	1780	9	HSA308251	AJ308251 Homo sapi
9	1729.4	50.1	219574	9	AP002956	AP002956 Homo sapi
10	1721	49.8	190762	2	AP000833	AP000833 Homo sapi
11	1717.4	49.7	144794	2	AP001182	AP001182 Homo sapi
12	1683.8	48.7	123159	2	HSAC000384	AC000384 Homo sapi
13	1633	47.3	2056	10	AY040865	AY040865 Mus muscu
14	1554.8	45.0	2730	10	BC023077	BC023077 Mus muscu
15	1042.2	30.2	2287	10	MMABC8MR	Z48745 M.musculus
16	1042.2	30.2	3586	6	AX305503	AX305503 Sequence
17	1042.2	30.2	3586	10	MMU34920	U34920 Mus musculus
18	1042.2	30.2	3624	10	AF323659	AF323659 Mus muscu
19	1025.2	29.7	3527	10	RNO303374	AJ303374 Rattus no
20	1015.2	29.4	2996	9	BC029158	BC029158 Homo sapi
21	1012	29.3	2805	9	AF323663	AF323663 Homo sapi
22	1012	29.3	2946	9	AF323662	AF323662 Homo sapi
23	1012	29.3	2983	9	AF323664	AF323664 Homo sapi
24	1012	29.3	3060	9	AF323661	AF323661 Homo sapi
25	1012	29.3	3142	9	AF323660	AF323660 Homo sapi
26	972.8	28.2	2745	9	HSU34919	U34919 Human white
27	972.8	28.2	2930	6	AX331149	AX331149 Sequence
28	972.8	28.2	2930	9	HSWHITE	X91249 H.sapiens m
29	897.8	26.0	3498	9	AY048757	AY048757 Homo sapi
30	747.8	21.6	173657	2	AC112557	AC112557 Rattus no
31	692.8	20.1	249283	2	AC124577	AC124577 Mus muscu
32	669.4	19.4	1327	10	BC026477	BC026477 Mus muscu
33	618.2	17.9	2833	9	AK092959	AK092959 Homo sapi
34	533.2	15.4	687	6	AX430376	AX430376 Sequence
35	429.6	12.4	1497	9	AF038175	AF038175 Homo sapi
36	396.8	11.5	2372	3	AY052113	AY052113 Drosophil
37	324.2	9.4	2537	6	E30646	E30646 Silkworm Bm
38	316.8	9.2	4498	3	AB030450	AB030450 Drosophil
39	313.6	9.1	4184	3	AY119101	AY119101 Drosophil
40	248.4	7.2	2631	3	AY069863	AY069863 Drosophil
41	243.6	7.1	111023	3	AC092242	AC092242 Drosophil
42	243.6	7.1	132666	3	AC005270	AC005270 Drosophil
43	243.6	7.1	218794	2	AC018064	AC018064 Drosophil
44	243.6	7.1	304634	3	AE003576	AE003576 Drosophil
45	243	7.0	3414	3	AY069801	AY069801 Drosophil

ALIGNMENTS

RESULT 1
HSA308237
LOCUS
DEFINITION Homo sapiens mRNA for ABC transporter (ABCG4 gene).
ACCESSION AJ308237
VERSION AJ308237.1 GI:20135821
KEYWORDS ABC transporter; ABCG4 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3498)
AUTHORS Engel, T., Lorkowski, S., Lueken, A., Rust, S., Schlueter, B.,
Berger, G., Cullen, P. and Assmann, G.
TITLE The human ABCG4 gene is regulated by oxysterols and retinoids in

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE

monocyte-derived macrophages
Biochem. Biophys. Res. Commun. 288, 488-493 (2001)
2 (bases 1 to 3498)
Engel, T.
Direct Submission
Submitted (25-AUG-2001) Macrophage Metabolism, Institute of
Arteriosclerosis Research, Domagkstrasse 3, Muenster 48149, Germany
Location/Qualifiers
1. .3498
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .3498
/gene="ABC G4"
12. .1952
/gene="ABC G4"
/codon_start=1
/product="ABC transporter"
/protein_id="CAC87131.1"
/db_xref="GI:20135822"
/translation="MAEKALEAVGCGLGPVAVAMVTLDDGAEPVLTTHLKKVENHI
TEAQRSHLPKBSADVIEVELSVREGPCWRKRGYKTLKCLSKGFCRRELIGIM
PSGAKSTPMNLAGYRESGMKGQILVNGRPRELFRKMSYIMQDDMLPLHLVLE
AMVSNLKLSEKQVKRELVEITLALGMSCHRTALLSGGQRKRLAIALELVNN
PVMFDEFTSLDASCPQVSLKSLAQGGRTIETIQHPSAKLFEMFDKLYLSQ
GQCIFGVVTLNIPYLKGLGLHCPYHNPADFTIEVASGEYGLNPLMFLRVQNGLCA
MAEKSSPEKNEVPAPCPPEVDPIESHTFATSLTOFCILFKRTFLSLRDVLT
HLRFMSHVIVGLGLLHIGDDASKVENNTGCLFSLFLMFLMFAALMPTLITPLEM
AVFMREHLNYSKRAYILAKTMADVPFOVPCVYCSIVYWMTPGAPSTSRFLFSA
LYATAVAOSLGLLIGLAASNSLOVATFVGPVTAIPVLLFSGFFVSFTIPTYLWSS
LASYRYGEGVILTIYGMERGLDLCLEERCPRFEPQSLRALDVEDAKLYNDLFLVG
IFFLALRLIAYLVLRVYKSR"

BASE COUNT 654 a 1098 c 934 g 812 t
ORIGIN

Query Match 99.2%; Score 3426.6; DB 9; Length 3498;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 3435; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY	7	ATGGCGGAGAGCGCTGAGCGCGTGGGCTGTGGACTAGGCGCGGGGCTGTGGCCATG	66
DB	12	ATGGCGGAGAGCGCTGAGCGCGTGGGCTGTGGACTAGGCGCGGGGCTGTGGCCATG	71
QY	67	GCCGTGACCTGTGAGGACGGGCGGAACCCCTGTGCTGACACGACCCCTGAAGAAGTG	126
DB	72	GCGGTGACCTGTGAGGACGGGCGGAACCCCTGTGCTGACACGACCCCTGAAGAAGTG	131
QY	127	GAGACCATCACTGAGACCCAGCGCTTCTCCACCTGCCAAGCGCTACGCGTGGAC	186
DB	132	GAGACCATCACTGAGACCCAGCGCTTCTCCACCTGCCAAGCGCTACGCGTGGAC	191
QY	187	ATCGAGTTCGTGAGCTGTCTATTCCGTCGGGAGGGGCGCTGTGGCGCAAAAGGGT	246
DB	192	ATCGAGTTCGTGAGCTGTCTATTCCGTCGGGAGGGGCGCTGTGGCGCAAAAGGGT	251
QY	247	TATAAGACCTTCTCAAGTGCCTCTCAGGTAAATTTGCGCGGGAGGCTGATGGCATC	306
DB	252	TATAAGACCTTCTCAAGTGCCTCTCAGGTAAATTTGCGCGGGAGGCTGATGGCATC	311
QY	307	ATGGGCCCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG	366
DB	312	ATGGGCCCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG	371
QY	367	TCTGGAATGAAGGGCAGATCTGGTTAATGAAGGCCACGGAGCTGAGGACCTTCGCG	426
DB	372	TCTGGAATGAAGGGCAGATCTGGTTAATGAAGGCCACGGAGCTGAGGACCTTCGCG	431
QY	427	AAAGTGTCTGTACATCATGCAAGATGACATGCTGCTGCCACACCTCACGGTGTGGAA	486
DB	432	AAAGTGTCTGTACATCATGCAAGATGACATGCTGCTGCCACACCTCACGGTGTGGAA	491
QY	487	GCCATGATGGTCTCTGCTAACTGAACTTACTGAGATCCCGATGTGAAAACGATCTC	546
DB	492	GCCATGATGGTCTCTGCTAACTGAACTTACTGAGATCCCGATGTGAAAACGATCTC	551

1627 ACCGCCATCCCTGCTCTCTGTTCTCCGGCTTCTTTGTGACGTTCAAGACCATCCCACT 1686
1632 ACCGCCATCCCTGCTCTCTGTTCTCCGGCTTCTTTGTGACGTTCAAGACCATCCCACT 1691
1687 TACCTGCAATGAGCTCTCTATCTCTCTATGTCAGGATGCTTTGAGGCTGTGATCTG 1746
1692 TACCTGCAATGAGCTCTCTATCTCTCTATGTCAGGATGCTTTGAGGCTGTGATCTG 1751
1747 ACGATCTATGCAATGAGCTCTCTATCTCTCTATGTCAGGATGCTTTGAGGCTGTGATCTG 1806
1752 ACGATCTATGCAATGAGCTCTCTATCTCTCTATGTCAGGATGCTTTGAGGCTGTGATCTG 1811
1807 GAGCACAGAGATCCTCCGAGCGCTGATGTCAGGATGCTTTGAGGCTGTGATCTG 1866
1812 GAGCACAGAGATCCTCCGAGCGCTGATGTCAGGATGCTTTGAGGCTGTGATCTG 1871
1867 CTGCTGTTGGGATCTCTCTCTAGCTCTCCGCTGCTGGCTACCTTGTGCTGCTGTTAC 1926
1872 CTGCTGTTGGGATCTCTCTCTAGCTCTCCGCTGCTGGCTACCTTGTGCTGCTGTTAC 1931
1927 CGGTCAGATCAGAGATAGAGGCTTGCCCGAGCTGTACCCAGCCCTGTCAGCAGGA 1986
1932 CGGTCAGATCAGAGATAGAGGCTTGCCCGAGCTGTACCCAGCCCTGTCAGCAGGA 1991
1987 AGCCCCAGTCCAGCCCTTTGGGACTGTTTAACTTATAGACTTGGGCTGCTGCT 2046
1992 AGCCCCAGTCCAGCCCTTTGGGACTGTTTAACTTATAGACTTGGGCTGCTGCT 2051
2047 GCGGGGCTATCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2106
2052 GCGGGGCTATCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2111
2107 AGCTGGGCTCTGGGAGTGGGCTTCCAGCCCTTCCAGCTATGCCCAGGAGTCTTCCCAA 2166
2112 AGCTGGGCTCTGGGAGTGGGCTTCCAGCCCTTCCAGCTATGCCCAGGAGTCTTCCCAA 2171
2167 GTTGATGGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2226
2172 GTTGATGGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2231
2227 GGAGCTGCTGCT 2286
2232 GGAGCTGCTGCT 2291
2287 AGGCTCTCTAGGCCCCACTTACAACTGACCAAGTGGCCCCCTCTGGGGTCCCAACA 2346
2292 AAGCTCTCTAGGCCCCACTTACAACTGACCAAGTGGCCCCCTCTGGGGTCCCAACA 2351
2347 CACAAGTGTGTAACCTGGGCTGCTATAAGGTTGGAGTTCCAGGCTGGGCTGGTGG 2406
2352 CACAAGTGTGTAACCTGGGCTGCTATAAGGTTGGAGTTCCAGGCTGGGCTGGTGG 2411
2407 AGTCCACTGGAATGCCATATGATGTTGAATGACAGGAGGAGTCTGGAAGTCTC 2466
2412 AGTCCACTGGAATGCCATATGATGTTGAATGACAGGAGGAGTCTGGAAGTCTC 2471
2467 TTCCT 2526
2472 TTCCT 2531
2527 ACAGAAGCTGGGTTTCTGCTAGTACCACTCCCAATCTCTGGGGATTTGGAGGCTG 2586
2532 ACAGAAGCTGGGTTTCTGCTAGTACCACTCCCAATCTCTGGGGATTTGGAGGCTG 2591
2587 GGGCTGGGATGCCATCCCTCCCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2646
2592 GGGCTGGGATGCCATCCCTCCCAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2651
2647 ACCTGTGCAATATCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2706
2652 ACCTGTGCAATATCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2711
2707 TATCTGGGCGGGGCTGAGTGGGGAAGACCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 2766

2712 TATCTGGGCGGGGCTGAGTGGGGAAGACCAACCTCTCTCTCTCTCTCTCTCTCTCTCT 2771
2767 CATGCACGCTCTGCTGCT 2826
2772 CATGCACGCTCTGCTGCT 2831
2827 GGCATCTCAGCCCTACACATTTGCCATCCCTACAGCACAGAGAAAGTGTGATGCTG 2886
2832 GGCATCTCAGCCCTACACATTTGCCATCCCTACAGCACAGAGAAAGTGTGATGCTG 2891
2887 ATGCTGCTGGTGGGCTGCTGCTGGGAGGACAGTGCACCTCTCTCTCTCTCTCTCTCT 2946
2892 ATGCTGCTGGTGGGCTGCTGCTGGGAGGACAGTGCACCTCTCTCTCTCTCTCTCTCT 2951
2947 GTTGAGGACTCTAAGATAAGCTGCTGCTGCCCGAGGCTGTCTACAGAACTGCAAGTCT 3006
2952 GTTGAGGACTCTAAGATAAGCTGCTGCTGCCCGAGGCTGTCTACAGAACTGCAAGTCT 3011
3007 CTACCCCCAAGCTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3066
3012 CTACCCCCAAGCTCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3071
3067 GTTCAGGAACCAACACAGCACACACCGGGCATAGTTGGCTTGGCCACTGCCACCA 3126
3072 GTTCAGGAACCAACACAGCACACACCGGGCATAGTTGGCTTGGCCACTGCCACCA 3131
3127 CGGCT 3186
3132 CGGCT 3191
3187 GCTGCTCATTAACCT 3246
3192 GCTGCTCATTAACCT 3251
3247 GCATCAGAACAGCTGCCCTGGGACACAGGTTGGGACACACTCAGAGCATGTCTGGCT 3306
3252 GCATCAGAACAGCTGCCCTGGGACACAGGTTGGGACACACTCAGAGCATGTCTGGCT 3311
3307 TTCTGCTGGTGGGCTCAGGCTCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3366
3312 TTCTGCTGGTGGGCTCAGGCTCATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3371
3367 CCCTTTTCT 3426
3372 CCCTTTTCT 3431
3427 CACACAGGATGCCAGTTGATTTGTTGGG 3455
3432 CACACAGGATGCCAGTTGATTTGTTGGG 3460

RESULT 2
HSA300465
LOCUS
DEFINITION
HSA300465
Homo sapiens mRNA for putative white family ATP-binding cassette transporter (ABCG4 gene).
ACCESSION
AJ300465
VERSION
AJ300465.1
KEYWORDS
GI:11342540
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Oldfield, S., Lowry, C.A. and Lightman, S.L.
TITLE
ABCG4: a novel white family ATP-binding cassette transporter expressed in human hypothalamus
JOURNAL
Abstr. - Soc. Neurosci. 27, 1951-1951 (2001)
REFERENCE
2 (bases 1 to 3501)
AUTHORS
Oldfield, S.
TITLE
Direct Submission
JOURNAL
Submitted (22-NOV-2000) Oldfield S., University Research Centre for

HSA300465 3501 bp mRNA linear PRI 13-MAR-2002
Homo sapiens mRNA for putative white family ATP-binding cassette transporter (ABCG4 gene).
AJ300465
AJ300465.1 GI:11342540
ABCG4 gene; white family ATP-binding cassette transporter.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Oldfield, S., Lowry, C.A. and Lightman, S.L.
ABCG4: a novel white family ATP-binding cassette transporter expressed in human hypothalamus
Abstr. - Soc. Neurosci. 27, 1951-1951 (2001)
2 (bases 1 to 3501)
Oldfield, S.
Direct Submission
Submitted (22-NOV-2000) Oldfield S., University Research Centre for

Neuroendocrinology, University of Bristol, Bristol royal Infirmary,
Marlborough Street, Bristol, BS2 8HW, UNITED KINGDOM

FEATURES
Source

1..3501
/organism="Homo sapiens"
/db_xref="RZPD:IMAGp998E133893"
/db_xref="taxon:9606"
/clone="I.M.A.G.E. consortium clone ID 1537140"
/tissue_type="Dorsal root ganglion"

gene

1..3501
/gene="ABCG4 gene"

5'UTR

1..56
/gene="ABCG4 gene"

CDS

57..1940
/gene="ABCG4 gene"

/codon_start=1
/product="putative white family ATP-binding cassette transporter"

/protein_id="CAC17140.1"

/db_xref="GI:11342541"

/db_xref="SWISS-PROT:O9H172"

/translation="NAVLEDDGAEPPVLTHLKKVENHITEAQRESHLPLKRSADVIEP
VELTSVREGPCWRKRGKTLKLSGKFCRRELIGIMPSPGAGKSTFNMILLAGYRES
GMKGILNGRPRLRTRFRKMSCYIMQDDMLPHLELVLEAMVSNLSEKQEVKKE
LVTEILALGLMSCSHRTALLSGQRRLAIALELVNPNVPMFDEPTSGIDSASF
QVSLMKLAOGGRTICTIHOPIAKLFEMFDKLYILSGOCIFKGVVNTLPIYKGL
GLHCPTYHNPADFIIEVAGSEYGDINPMLFRAVONGLCAMEKSSPEKNEVPAPCP
CPPEVDPIESHTFATSTORFELSLRDLTLHLRPMHSHVIVGULIGLLYL
HIGDDAKFVNNTGCLFESMLFMFAALPMTVLPFLEMAVFMREHLNINWYSLKAYLL
AKTMADVPVPCVPCVYCSIVYMTQGAETSRFLFSALATATLVAQSLGLLIGAA
SNSLQVATFEPVPTAIVLLFSGFVSKFTIPTLYLOWSSYLSYVRYGFEVILTIYGM
ERGDILTCLEERCFPEPOSILLRALDVEDAKLYMDFLVLGIFFLALRLAYLLVLRVK
SER"

3'UTR

1941..3501

/gene="ABCG4 gene"

BASE COUNT 668 a 1095 c 927 g 810 t 1 others

ORIGIN

Query Match 99.1%; Score 3424.6; DB 9; Length 3501;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 3433; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 8 TGGCGGAGAGCGCTGGAGGCGGTGGGCTGTGGACTAGGGCCGGGGCTGTGGCCATGG 67
DB 1 TGGCGGAGAGCGCTGGAGGCGGTGGGCTGTGGACTAGGGCCGGGGCTGTGGCCATGG 60

QY 68 CCGTGACGCTGGAGGACGGGGCGGAACCCCTGTCTGACACGACCACTGAAGAGGTGG 127
DB 61 CCGTGACGCTGGAGGACGGGGCGGAACCCCTGTCTGACACGACCACTGAAGAGGTGG 120

128 AGAACCATCACTCAAGCCAGCGCTTCTCCACCTGCCAAGCGCTCAGCGGTGGACA 187
DB 121 AGAACCATCACTCAAGCCAGCGCTTCTCCACCTGCCAAGCGCTCAGCGGTGGACA 180

QY 188 TGGAGTTCTGGAGCTGCTTCCATTCCTGCGGGAGGGGCCCTGTGGCGCAAAAGGGTT 247
DB 181 TGGAGTTCTGGAGCTGCTTCCATTCCTGCGGGAGGGGCCCTGTGGCGCAAAAGGGTT 240

QY 248 ATAAGACCTTCTCAAGTGGCTCTCAGCTAAATTCGCGCGGGAGGCTGATGCGATCA 307
DB 241 ATAAGACCTTCTCAAGTGGCTCTCAGCTAAATTCGCGCGGGAGGCTGATGCGATCA 300

QY 308 TGGGCCCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGGAGT 367
DB 301 TGGGCCCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGGAGT 360

QY 368 CTGGAATGAAGGGGAGAGTCTGGTTAATGAAGGCCACGGAGCTGAGGACCTTCCGCA 427
DB 361 CTGGAATGAAGGGGAGAGTCTGGTTAATGAAGGCCACGGAGCTGAGGACCTTCCGCA 420

QY 428 AGATGCTCTGATCATATGAAGATGACATGCTGCTGCGGCACCTCAGCGTGTGGAAG 487
DB 421 AGATGCTCTGATCATATGAAGATGACATGCTGCTGCGGCACCTCAGCGTGTGGAAG 480

QY 488 CCATGATGCTCTGCTAACTGAATCTTACTGAGATCCCGATGTGAAAAAGCATCTCG 547
DB 481 CCATGATGCTCTGCTAACTGAAGCTGAGTGAAGCAGGAGTGAAGAAGGAGCTGG 540

QY 548 TGACAGAGATCTCGACGGCACTGGGCTGATGCTGCTGCCACACGAGGAGACGCCCTGC 607
DB 541 TGACAGAGATCTCGACGGCACTGGGCTGATGCTGCTGCCACACGAGGAGACGCCCTGC 600

QY 608 TCTCTGGGGGAGAGGAGGCTGTGGCCATCGCCCTGGAGCTGTGTTCAACACCCGCCCTG 667
DB 601 TCTCTGGGGGAGAGGAGGCTGTGGCCATCGCCCTGGAGCTGTGTTCAACACCCGCCCTG 660

QY 668 TCATGTTCTTTCATGAGCCACACAGTGGTCTGATGAGCCCTCTGTTTCCAAAGTGGTGT 727
DB 661 TCATGTTCTTTCATGAGCCACACAGTGGTCTGATGAGCCCTCTGTTTCCAAAGTGGTGT 720

QY 728 CCTCATGAAGTCCCTGGCACAGGGGGCGCTGACCATCATCTGCACCATCCACACGCCCA 787
DB 721 CCTCATGAAGTCCCTGGCACAGGGGGCGCTGACCATCATCTGCACCATCCACACGCCCA 780

QY 788 GTGCCAAGCTCTTTGAGATGTTTGACAAAGCTCTACATCTGAGCAGGGTCAAGTGCATCT 847
DB 781 GTGCCAAGCTCTTTGAGATGTTTGACAAAGCTCTACATCTGAGCAGGGTCAAGTGCATCT 840

QY 848 TCAAAAGGAGTGTTCACCAACCTGATCCCTATCTAAAGGAGCTCGGCTTGCATTTGCCCA 907
DB 841 TCAAAAGGAGTGTTCACCAACCTGATCCCTATCTAAAGGAGCTCGGCTTGCATTTGCCCA 900

QY 908 CCTACCAACACCGGCTGACTTTCATCATGAGGTGGGCTCTGGCAGTATGGAGACCTGA 967
DB 901 CCTACCAACACCGGCTGACTTTCATCATGAGGTGGGCTCTGGCAGTATGGAGACCTGA 960

QY 968 ACCCATGTTGTTGAGGCTGTGCAGATGGGCTGTGGCTATGGCTGAGAGAGAGCA 1027
DB 961 ACCCATGTTGTTGAGGCTGTGCAGATGGGCTGTGGCTATGGCTGAGAGAGAGCA 1020

QY 1028 GGCCTGAGAAGAACGAGGTCCCTGCCCATGCCCCTTGTCTTCCCGGAAGTGGATCCCA 1087
DB 1021 GGCCTGAGAAGAACGAGGTCCCTGCCCATGCCCCTTGTCTTCCCGGAAGTGGATCCCA 1080

QY 1088 TTGAAAGCCACACCTTTGCCACGACCCCTCACACAGTTCTGCATCTTCAAGAGCA 1147
DB 1081 TTGAAAGCCACACCTTTGCCACGACCCCTCACACAGTTCTGCATCTTCAAGAGCA 1140

QY 1148 CTTTCTCTTCCATCTCAGGACACGGTCTCTGACCCACCTAGGTTTCATGTCACAGTGG 1207
DB 1141 CTTTCTCTTCCATCTCAGGACACGGTCTCTGACCCACCTAGGTTTCATGTCACAGTGG 1200

QY 1208 TTATTGGGCTGTCTCATCGGCTTCTTACCTGCATATTGGCGACCATGCCAGAGTCT 1267
DB 1201 TTATTGGGCTGTCTCATCGGCTTCTTACCTGCATATTGGCGACCATGCCAGAGTCT 1260

QY 1268 TCAACAACACGGGCTGCTTCTTCTTCTCCATGCTGTTCTCATGTTGCGCGCCCTCATGC 1327
DB 1261 TCAACAACACGGGCTGCTTCTTCTTCTCCATGCTGTTCTCATGTTGCGCGCCCTCATGC 1320

QY 1328 CAACCTGCTCACCCTTCCCTTAGAGATGGCGGCTTTCATGAGGAGCACCTCAACTACT 1387
DB 1321 CAACCTGCTCACCCTTCCCTTAGAGATGGCGGCTTTCATGAGGAGCACCTCAACTACT 1380

QY 1388 GGTACAGCTCAAAGCGTATTACCTGGCCAAAGACCATGGCTGACCTGCCCTTTCAGGTGG 1447
DB 1381 GGTACAGCTCAAAGCGTATTACCTGGCCAAAGACCATGGCTGACCTGCCCTTTCAGGTGG 1440

QY 1448 TGTGTCGGTGGTCTACTGACAGTGTGTACTTGGATGACGGGCGAGCCGCTGAGACCA 1507
DB 1441 TGTGTCGGTGGTCTACTGACAGTGTGTACTTGGATGACGGGCGAGCCGCTGAGACCA 1500

QY 1508 GCCGCTTCTCTCTCAGCCCTGGCCACCGCCCTTGGTGGCCCAATCTTTGG 1567
DB 1501 GCCGCTTCTCTCTCAGCCCTGGCCACCGCCCTTGGTGGCCCAATCTTTGG 1560

QY 1568 GGCTGCTGATCGAGAGTCTTCCAACTCCCTACAGGTGGCCACTTTTGTGGGCCAGTTA 1627
Db 1561 GGTCTGATCGGAGTGTCTTCAACTCCCTACAGGTGGCCACTTTTGTGGGCCAGTTA 1620
QY 1628 CCAGCATCCCTGCTCTTCTTCTCGGCTTCTTGTGACGTTCAAGACCATCCCACTT 1687
Db 1621 CCAGCATCCCTGCTCTTCTTCTCGGCTTCTTGTGACGTTCAAGACCATCCCACTT 1680
QY 1688 ACCTGCAATGGAGTCCCTATCTCTCTATGTCAGGTATGGCTTGTAGGGTGTGATCTGA 1747
Db 1681 ACCTGCAATGGAGTCCCTATCTCTCTATGTCAGGTATGGCTTGTAGGGTGTGATCTGA 1740
QY 1748 CGATCTATGATCGAGGAGAGACCTGACATGTTTATAGAGAACGCTGCCGTTCCGGG 1807
Db 1741 CGATCTATGATCGAGGAGAGACCTGACATGTTTATAGAGAACGCTGCCGTTCCGGG 1800
QY 1808 AGCCACAGACATCCCTCCGAGCGCTGGATCTGAGGATGCCAAGCTCTACATGCACTTC 1867
Db 1801 AGCCACAGACATCCCTCCGAGCGCTGGATCTGAGGATGCCAAGCTCTACATGCACTTC 1860
QY 1868 TGGTCTTGGGCACTTCTTCTTCTAGCCCTGGGCTGTGCTGGCTTACCTTGTCTGCTTTACC 1927
Db 1861 TGGTCTTGGGCACTTCTTCTTCTAGCCCTGGGCTGTGCTGGCTTACCTTGTCTGCTTTACC 1920
QY 1928 GGTCTAAGTCAGAGAGATAGAGGCTTGCCTCCAGCGCTGTACCCAGCCCTGCGAGAGAA 1987
Db 1921 GGTCTAAGTCAGAGAGATAGAGGCTTGCCTCCAGCGCTGTACCCAGCCCTGCGAGAGAA 1980
QY 1988 GCCCCAGTCCAGCCCTTGGGACTGTTTAACTTTATAGACTTGGGCACTGGTTCCTG 2047
Db 1981 GCCCCAGTCCAGCCCTTGGGACTGTTTAACTTTATAGACTTGGGCACTGGTTCCTG 2040
QY 2048 GCGGGGCTATCTCTCTCTCTTGGCTTCCACAGGCTGGCTGGCTGGCTCCCA 2107
Db 2041 GCGGGGCTATCTCTCTCTCTCTTGGCTTCCACAGGCTGGCTGGCTGGCTCCCA 2100
QY 2108 GCCTGGGCTGGGAGTGGGGCTTCCAGCCCTCCCACTATGCCAGGAGTCTTCCCAAG 2167
Db 2101 GCCTGGGCTGGGAGTGGGGCTTCCAGCCCTCCCACTATGCCAGGAGTCTTCCCAAG 2160
QY 2168 TTGATGGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2227
Db 2161 TTGATGGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
QY 2228 GAGGCTGCTGCCT 2287
Db 2221 GAGGCTGCTGCCT 2280
QY 2288 GGCTCTCTAGGCCCCACTTACAACTGACCAAGTGGCCCCCTCTGGGGGTGCCCAACC 2347
Db 2281 AGCTCTCTAGGCCCCACTTACAACTGACCAAGTGGCCCCCTCTGGGGGTGCCCAACC 2340
QY 2348 ACAAGTCTTGTAACTGGGCTGCTATAAGTGTGGAGTTCACAGGCTGGGCCCTGGTGA 2407
Db 2341 ACAAGTCTTGTAACTGGGCTGCTATAAGTGTGGAGTTCACAGGCTGGGCCCTGGTGA 2400
QY 2408 GTCCACTGGAAGTCCCAATTATGATGTTGAAATGGACAGGGAAGGACTCTGGAAGTCTCT 2467
Db 2401 GTCCACTGGAAGTCCCAATTATGATGTTGAAATGGACAGGGAAGGACTCTGGAAGTCTCT 2460
QY 2468 TCCT 2527
Db 2461 TCCT 2520
QY 2528 CAGAAGCTGGGTTTCTGTCTAGTCACTACCTCCCAATCTCTGGGATGGAGAGGCTGG 2587
Db 2521 CAGAAGCTGGGTTTCTGTCTAGTCACTACCTCCCAATCTCTGGGATGGAGAGGCTGG 2580
QY 2588 GGCTGGGATGCCCACTCCCTCTCCCACTACCTCTTGTGGGGGAGGCTGGTGGCA 2647
Db 2581 GGCTGGGATGCCCACTCCCTCTCCCACTACCTCTTGTGGGGGAGGCTGGTGGCA 2640
QY 2648 CCGTGTGCAATATGCTGTGTTCTCTCCCACTGGCACTGGAAATGCACTTTT 2707

Db 2641 CCTGTGCAATATGCTGTGTTCTCTCCCACTGGCACTGGAGATGCACTTT 2700
QY 2708 ATTCTGGGGGGGTGAGTGGGGAAGACCCAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2767
Db 2701 ATTCTGGGGGGGTGAGTGGGGAAGACCCAACTCTCTCTCTCTCTCTCTCTCTCTCTCT 2760
QY 2768 ATGACAGGCTCTCGTATGCT 2827
Db 2761 ATGACAGGCTCTCGTATGCT 2820
QY 2828 GCCATCTCAGCCCTACACACTTCCATCTCCCTCTACAGCACAGAGAGTGTGGTGGCA 2887
Db 2821 GCCATCTCAGCCCTACACACTTCCATCTCCCTCTACAGCACAGAGAGTGTGGTGGCA 2880
QY 2888 TGTCTGTGTGGGGGTGCTGGTGGGAGGACGTCACCACTCTCTCTCTCTCTCTCTCTCTCT 2947
Db 2881 TGTCTGTGTGGGGGTGCTGGTGGGAGGACGTCACCACTCTCTCTCTCTCTCTCTCTCTCT 2940
QY 2948 TTGGAGACTCTAAGGATAAGGCTGTCTGCCAGGCTGTCTACAGAACTGCAGGTGTC 3007
Db 2941 TTGGAGACTCTAAGGATAAGGCTGTCTGCCAGGCTGTCTACAGAACTGCAGGTGTC 3000
QY 3008 TACCCCAAGTCTTCT 3067
Db 3001 TACCCCAAGTCTTCT 3060
QY 3068 TTGAGGAACCAACACAGACCAACACCGGCACTAAGTTGGCTTGGCCACTGCCACCCAC 3127
Db 3061 TTGAGGAACCAACACAGACCAACACCGGCACTAAGTTGGCTTGGCCACTGCCACCCAC 3120
QY 3128 GGCCCT 3187
Db 3121 GGCCCT 3180
QY 3188 CTGCTCATTCAAACTTCTGTCCATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3247
Db 3181 CTGCTCATTCAAACTTCTGTCCATGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3240
QY 3248 GCATCAGAACAGCTGCCCTGGCCACAGGTGGCAGACACACTCAGAGCATGTCTGGCTT 3307
Db 3241 GCATCAGAACAGCTGCCCTGGCCACAGGTGGCAGACACACTCAGAGCATGTCTGGCTT 3300
QY 3308 TCCTGTGGGTCCAGGCTCATCT 3367
Db 3301 TCCTGTGGGTCCAGGCTCATCT 3360
QY 3368 CTTTCT 3427
Db 3361 CTTTCT 3420
QY 3428 ACACAGGATGCCAGTGTGTTTGGG 3455.
Db 3421 ACACAGGATGCCAGTGTGTTTGGG 3448

RESULT 3

HSM802306

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

HSM802306 2838 bp mRNA linear PRI 18-FEB-2000
Homo sapiens mRNA; cDNA DKF2p434p1420 (from clone DKF2p434p1420);
partial cds.

AL137563
AL137563.1 GI:6808269

Homo sapiens.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2838)
Ottenwaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and
Wiemann, S.

Direct Submission
Submitted (15-JAN-2000) MIPS, Am Klopferspitz 18a, D-82152

Martinsried, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by Medigenomix (Martinsried/Germany) within the cDNA
 sequencing consortium of the German Genome Project. This clone
 (DKFZp434P1420) is available at the RZPD in Berlin. Please contact
 the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
 information about the clone and the sequencing project is available
 at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES

source

1. .2838
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="11"
 /clone="DKFZp434P1420"
 /tissue_type="testis"
 /clone_lib="434 (synonym: htes3). Vector pSport1; host
 DH10B; sites NotI + SalI"
 /dev_stage="adult"
 1. .1253
 /gene="DKFZp434P1420"
 <1. .1253
 /gene="DKFZp434P1420"
 /note="strong similarity to mouse white protein"
 /codon_start=3
 /product="hypothetical protein"
 /protein_id="CA870814.1"
 /db_xref="GI:6808270"
 /translation="LDSASCFQVVSIMLSLAQGRITICTIHOPSAKLFEFMDKLYIL
 SOQCIFKGVNLPYILKGLGHCTPHNPADFIIEVASGEYGLNPLMFRAVQGL
 CMAEKSSPEKNEVPAPCPPEVDPIESHFTATSTLQFCILPKRFLSLRDTV
 LTHLRMSHWVIGVLGLYLHIGDDASKVENNTGCLFFSMLFLMFAALMPVLPTPL
 EMAVENHLNWSLYLKAYLAKTMADVPFQVPCVYGSYIYWMGTQPAETSRFLFL
 SALATATVAOSLGLGIAANSLSQADVFVGPVTAIPVLFSGFVSEKTIPTYLQW
 SSYLSVRYGFEVILTIYMERGDLTCLRECPFPREPOSILRALDVEDAKLYMDFLV
 LGIFFLRLRLYLRLRYVKSER"
 2777. .2782
 polyA_signal
 polyA_site
 2796

BASE COUNT 548 a 909 c 702 g 679 t

ORIGIN

Query Match 79.9%; Score 2761; DB 9; Length 2838;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 695 GTCTGATAGCGCTCTTGTTCCTCAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGGGG 754
 1 GTCTGATAGCGCTCTTGTTCCTCAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGGGG 60
 755 GCGGTACCATCATCTGCACCATCCACAGCCAGTCCCAAGCTCTTTGAGATCTTTGACA 814
 61 GCGGTACCATCATCTGCACCATCCACAGCCAGTCCCAAGCTCTTTGAGATCTTTGACA 120
 815 AGCTTACATCTCTGAGCCAGGCTAGTGCATCTTCAAAGGAGTGGTCAACCACTGATCC 874
 121 AGCTTACATCTCTGAGCCAGGCTAGTGCATCTTCAAAGGAGTGGTCAACCACTGATCC 180
 875 CCTATCTAAGGACTTCGCTTGCATTTGCCACCTTACCACACCGCGCTGATCTCATCA 934
 181 CCTATCTAAGGACTTCGCTTGCATTTGCCACCTTACCACACCGCGCTGATCTCATCA 240
 935 TCGAGTGGCCCTCTGGCGAGTATGGAGACTGAACCCCATGTGTTCAGGGCTGTGCAGA 994
 241 TCGAGTGGCCCTCTGGCGAGTATGGAGACTGAACCCCATGTGTTCAGGGCTGTGCAGA 300
 995 ATGGGCTGTGGCTATGGCTGAGAAGAAGAGCAGCCCTGAGAAGAAGAGGTCCTTGGCC 1054
 301 ATGGGCTGTGGCTATGGCTGAGAAGAAGAGCAGCCCTGAGAAGAAGAGGTCCTTGGCC 360
 1055 CATGCCCTCTCTCTCTCGGAGTGGATCCCATTTGAAGCCACACCTTTGGCACCACGA 1114
 361 CATGCCCTCTCTCTCTCGGAGTGGATCCCATTTGAAGCCACACCTTTGGCACCACGA 420

1115 CCCTCACACAGTTCTGCATCTCTTCAAGAGGACCTTCTGTCTCATCTCTCAGGACACGG 1174
 421 CCCTCACACAGTTCTGCATCTCTTCAAGAGGACCTTCTGTCTCATCTCTCAGGACACGG 480
 1175 TCCTGACCCACCTTACAGTGTTCATGCTCCACAGTGGTATTTGGCGTGTCTATCGGCTCTCT 1234
 481 TCCTGACCCACCTTACAGTGTTCATGCTCCACAGTGGTATTTGGCGTGTCTATCGGCTCTCT 540
 1235 ACCTGATATTTGGCGAGGATGCCAGCAAGTCTTCAACACACCGGCTGCTCTTCTTCT 1294
 541 ACCTGATATTTGGCGAGGATGCCAGCAAGTCTTCAACACACCGGCTGCTCTTCTTCT 600
 1295 CCATGCTGTTCCTCATGTTTCGGCGCTCATGCTCAACTGTGCTACCTTCCCTTTAGAGA 1354
 601 CCATGCTGTTCCTCATGTTTCGGCGCTCATGCTCAACTGTGCTACCTTCCCTTTAGAGA 660
 1355 TGCGGTCTTTCATGAGGAGACCTCAACTACTTGTACAGCTCAAGAGGTATTAACCTGG 1414
 661 TGCGGTCTTTCATGAGGAGACCTCAACTACTTGTACAGCTCAAGAGGTATTAACCTGG 720
 1415 CCAAGACCATGGCTGAGTGGCTTTCAGTGTGTGTCGGTGTCTACTGAGCATTTG 1474
 721 CCAAGACCATGGCTGAGTGGCTTTCAGTGTGTGTCGGTGTCTACTGAGCATTTG 780
 1475 TGCTACTGGATGAGGCGCAGCGCTGAGACGAGCGCTTCTGCTCTTCTCAGCCCTGG 1534
 781 TGCTACTGGATGAGGCGCAGCGCTGAGACGAGCGCTTCTGCTCTTCTCAGCCCTGG 840
 1535 CCACCCGCCCGCTTGGTGGCCCAATCTTTGGGGTGTGTCGAGCGAGCTGTCTTCAACT 1594
 841 CCACCCGCCCGCTTGGTGGCCCAATCTTTGGGGTGTGTCGAGCGAGCTGTCTTCAACT 900
 1595 CCCTACAGTGGCCACTTTTGGGGCCAGTTACCGCCATCCCTGCTCTTGTCTTCTCGG 1654
 901 CCCTACAGTGGCCACTTTTGGGGCCAGTTACCGCCATCCCTGCTCTTGTCTTCTCGG 960
 1655 GCTTCTTGTGCTAGCTTCAAGACCATCCCACTTACCTGCAATGGAGCTCTCTATCTCTCT 1714
 961 GCTTCTTGTGCTAGCTTCAAGACCATCCCACTTACCTGCAATGGAGCTCTCTATCTCTCT 1020
 1715 ATGTCAGTATGGCTTGGGGTGTGATCTCTACGATCTATGGCATGAGCGAGGAGACC 1774
 1021 ATGTCAGTATGGCTTGGGGTGTGATCTCTGAGGATCTATGGCATGAGCGAGGAGACC 1080
 1775 TGACATGTTTAGAGGAACCTGCCGTTCCGGGAGCCACAGACATCTCTCCGAGCGCTGG 1834
 1081 TGACATGTTTAGAGGAACCTGCCGTTCCGGGAGCCACAGACATCTCTCCGAGCGCTGG 1140
 1835 ATGTGGAGGATGCCAAGCTCTACATGAGCTTCTGCTGTGGCATCTTCTTCTTCTAGCCC 1894
 1141 ATGTGGAGGATGCCAAGCTCTACATGAGCTTCTGCTGTGGCATCTTCTTCTTCTAGCCC 1200
 1895 TCGGGTGTGGCTTACCTTGTGCTGCTTACCGGTCAAGTCAGAGATAGAGCTTTG 1954
 1201 TCGGGTGTGGCTTACCTTGTGCTGCTTACCGGTCAAGTCAGAGATAGAGCTTTG 1260
 1955 CCCCAGCCTGTACCCAGCCCTTGAGCAGAGAAAGCCCCAGTCCAGAGCCCTTTGGAGCTG 2014
 1261 CCCCAGCCTGTACCCAGCCCTTGAGCAGAGAAAGCCCCAGTCCAGAGCCCTTTGGAGCTG 1320
 2015 TTTTAACCTTATAGACTTGGGCACTGTTCTCTGGGGGCTATCTCTCTCTCTTCTTGGCT 2074
 1321 TTTTAACCTTATAGACTTGGGCACTGTTCTCTGGGGGCTATCTCTCTCTCTTCTTGGCT 1380
 2075 CTCCACAGGCTGGCTGTCGGACTGCGCTCCAGCCTGGGCTCTGGAGTGGGGCTCCA 2134
 1381 CTCCACAGGCTGGCTGTCGGACTGCGCTCCAGCCTGGGCTCTGGAGTGGGGCTCCA 1440
 2135 GCGCTCCCACTATGCCAGGAGTCTTCCCAAGTTGATCGGTTGTAGCTTCTTCCCTCA 2194
 1441 GCGCTCCCACTATGCCAGGAGTCTTCCCAAGTTGATCGGTTGTAGCTTCTTCCCTCA 1500

[illegible][illegible]

[illegible]

Db	3279	CTGTGGACATCAGAGCGCGGGGCTGTGACATCAGGTGGCAGACACAC	-----CIVGTC 3333
QY	3303	GGCTTTCTCTGGTGGGTCCAGGCTCATCTGCTTCTGATTTTCCCTCCCCAGGCGCTCATTT	3362
Db	3331	TATCTTCTTGGTGGGTCCAGGATCATCTATATCTGTTTCTCCTCCTCAGGACTCTTT	3390
QY	3363	TTCCCCCTTTTCTGTACACATCCCTGCTTACCTCCTCTACCCGCCACAGATCTTTC	3422
Db	3391	CCC-----CAGTCCCTGTGCTCACTCCTCTGTATCTGCCCCAGGTTCTTC	3435
QY	3423	CTATCACACAGGATGCCAGTGTGATTGTGGG	3455
Db	3436	TCATTACACAGAGA--CCAGTTGATTGTGGG	3466
RESULT 5			
LOCUS	BC016200	3522 bp mRNA linear	ROD 07-AUG-2002
DEFINITION	Mus musculus, Similar to ATP-binding cassette, sub-family G (WHITE), member 4, clone MGC:27647 IMAGE:4511456, mRNA, complete cds.		
ACCESSION	BC016200		
VERSION	BC016200.1	GI:16740636	
KEYWORDS	MGC.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 3522)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission		
	Submitted (29-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabbs-r@mail.nih.gov Tissue Procurement: The Cepko Laboratory cDNA Library Preparation: Life Technologies, Inc. cDNA Library: Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Varon Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lilisa Prabhu, Parvaneh Saedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAC Plate: 38 Row: k Column: 8 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein. Location/Qualifiers 1..3522 /organism="Mus musculus" /db_xref="taxon:10090" /clone="MGC:27647 IMAGE:4511456" /tissue_type="Eye, retina, mouse strain C57Bl/6" /clone_lib="NIH_MGC_94" /lab_host="DH10B" /note="vector: pCMV-SPORT6" 181..2064 /codon_start=1 /product="Similar to ATP-binding cassette, sub-family G (WHITE), member 4" /protein_id="AAH16200.1"		
CDS			

/db_xref="GI:16740637"
translation="MAVLEDDGAPVPLTTHLKKVENHTEAQRFSHLPKRSAYDIEF
VELSVREGPCWKRKYKTKLLKLSGKFCRELIGIMPGSCAKSKFMNLLAGYRES
CMKQILVNGPRELPRFKMSCYIMQDDMLLPHLVLEAMVWVANKLSKQEVKKE
LVTELLTALGMSCHRTALDGGQRRLAILELVNPPVMEFFDEPTSLDASCF
QVSLMKSLAGRTVICTIHOPSAKIFEMFDKLYILSQGCIKFGVYVNLIPYLKGL
GLHPTVHPADFTIEVASGEYDINPLFRAVONGICTMAEKSSKPKNELPAHCP
CPPELDTESHTFATSLTOPCILFRFTFLSILDRVTLTHURFMSHVLIGVLGLYL
HIGDDASKVENNTGFLFESMLFLMFAALMPTVLPFPEMAVEMREHLNRYWTLKAYIL
KTMADVPFQVQPVVCSIVYMTGQPAETSRLLFSALALATVAQSLGLIGAA
STSLQVTFVGPVTAIPVLFDSGFVSEFTIPTLYLOWSSYLSYVRYGFEGLITYGM
ERHGLTCLDEQCPRPQIILRELDEVEAKLYMDLVLGIFFLALRLAYLVLRVYK
SER"

BASE COUNT 675 a 1079 c 918 g 850 t

ORIGIN

Query Match 61.7%; Score 2132.2; DB 10; Length 3522;

Best Local Similarity 80.3%; Pred. No. 0;

Matches 2774; Conservative 0; Mismatches 568; Indels 111; Gaps 19;

Qy	7	ATGGCGGAGAAGCGCTGGAGCGCTGGCGTGTGGACTAGGCGGGGGCTGTGGCCATG	66
	124	ATGGCGGAGAAGCGATTGGAGCGCTGGCGTGTGGACTGGCGCTGGGGCTGTGGCCATG	183
Qy	67	GCGGTGACGCTGGAGGAGCGGGCGGAACCCCTGTGCTGACCGACCTCAAGAAAGTG	126
Db	184	GCTGTGACACTGGAGGACGGGCGAGACCCCTGTGCTGACCGACCTTAAGAGGTG	243
Qy	127	GAGAACACATCACTGAAGCCAGCGCTTCCACCTGCCCAAGCGCTAGCGGTGGAC	186
Db	244	GAGAACCATATCAGAAAGACAGCGCTTCTCACCTGCCCAAGCGCTCTGCTGGAC	303
Qy	187	ATCAGGTCGTGGAGCTGCTGCTATTCCGTGGGAGGGCGCTGTCGCGCAAAAGGCT	246
Db	304	ATCAGGTCGTGGAACTGCTCTATTCCGTGGGAGGGCGCTGTCGCGCAAAAGGCT	363
Qy	247	TATAAGACCTTCTCAAGTGCCTCTCAGGTAAATTTGCCCGCGGAGCTGATTGGCATC	306
Db	364	TACAAGACCTTCTCAAGTGCCTCTGCTGAAGTCTGTCGCGGAGAGCTGATTGGCATC	423
Qy	307	ATGGCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG	366
Db	424	ATGGCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG	483
Qy	367	TCTGAATGAAGGCGACATCTGGTTAATGAAGGCCAGCGGAGCTGAGGACCTTCGCG	426
Db	484	TCGGGATGAAGGCGAGATCCTGGTTAACGGGAGCGCGGAGCTGAGGACCTTCGCG	543
Qy	427	AAGATGCTCTCATCATCAAGATGACATGCTGCTGCCGACCTCACCGGTGTGGAA	486
	544	AAGATGCTCTCATCATCAAGATGACATGCTGCTGCCGACCTCACAGTCTCTGGAG	603
Qy	487	GCCATGATGCTCTGCTAACTGAATCTTACTGAGATCCCGATGTAAGAAAGATCTC	546
Db	604	GCCATGATGCTCTGCTAACTGAATCTTACTGAGATCCCGATGTAAGAAAGATCTC	663
Qy	547	GTGACAGATCTCTGACGGCATCTGGGCTGATGCTGCTCCACACAGGACAGCCCTG	606
Db	664	GTGACAGATCTCTGACGGCATCTGGGCTGATGCTGCTCCACACAGGACAGCTCTG	723
Qy	607	CTCTCTGGCGGCGAGAGAGCGTCTGGCCATCCGCTGGAGCTGGTGAACAAACCCGCT	666
Db	724	CTCTCTGGCGGCGAGAGAGCGCTGGCCATCTGCTGGAACCTGGTCAACAAACCCGCT	783
Qy	667	GTCATGCTTTGATGAGCCCAAGCGTCTGGATGCTGATGAGCGCTCTGTTTCCAGTGGT	726
Db	784	GTCATGCTTTGATGAGCCCAAGCGTCTGGATGCTGATGAGCGCTCTGTTTCCAGTGGT	843
Qy	727	TCCCTCATGAAGTCCCTGGGACAGGGGGCGGTACCATCATCTGCACCATCCACAGCC	786
Db	844	TCCCTCATGAAGTCCCTGGGACAGGGGGCGGTACCATCATCTGCACCATCCACAGCC	903
Qy	787	AGTGCCAAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTCTGAGCCAGGGTCACTGCATC	846

Db	904	AGTGCCAAAGCTCTTTGAGATCTTTGACAAGCTCTACATCTCTGAGCCAGGGCAATGCATC	963
Qy	847	TTCAAGAGAGTGTGCACCAACCTGATCCCTATCTAAAGGACATCGGCTTGATCTGATCC	906
Db	964	TTCAAGGCGGTGGTTACCAACCTGATCCCTATCTGAAGGGGCTTGGCTTGACATGTC	1023
Qy	907	ACCTACCAACAACCGGCTGACTTCTCATCAGAGTGGCTCTCGGAGTATGGAGACCTG	966
Db	1024	ACCTACCAACAACCGGCTGACTTCTCATCAGAGTGGCTCTCGGAGTATGGAGACCTG	1083
Qy	967	AACCCCATGTTTTCAGGGCTGTGAGAATGGGCTGTGCGCTATGGCTGAGAAGAAGAGC	1026
Db	1084	AACCCCATGTTTTCAGGGCTGTGAGAATGGGCTGTGCGCTATGGCTGAGAAGAAGAGC	1143
Qy	1027	AGCCCTGAGAAGAAGAGGCTCCCTGCGCCATGCGCCCTCTTGTCTCCGGAAGTGGATCC	1086
Db	1144	AGCCCTGAGAAGAATGAGTCCCTGCGCACCTGCGCCACTTGGCTCGGGAGCTGATCCT	1203
Qy	1087	ATTGAAGCCACACCTTTGCCACGACGACCTCACACAGTTCATCATCTCTTCAAGAGG	1146
Db	1204	ATTGAAGCCACACATTTGCCACGACGACCTAACCCAGTTCATCATCTCTTCAAGAGG	1263
Qy	1147	ACCTTCTCTGCTCCATCTCAGGGACACGCTCCTGACCCACCTTACGCTTCTATGCTCCAC	1206
Db	1264	ACCTTCTCTGCTCCATCTCAGGGACACGCTCCTGACCCACTTTCGCTTCTATGCTCAT	1323
Qy	1207	GTTATTTGGCTGTCATCGGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1266
Db	1324	TTGATCGGGCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1383
Qy	1267	TTCAACAAACCGGCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1326
Db	1384	TTCAACAAACCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1443
Qy	1327	CCAACTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1386
Db	1444	CCGACCGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1503
Qy	1387	TGGTACAGCTTCAAGCGTATTACTTGGCCAAAGACCATGGCTGACGTGCTTTCAGGTG	1446
Db	1504	TGGTATCTCTTCAAGCGTATTACTTGGCCAAAGACCATGGCTGATGTGCTTTCAGGTG	1563
Qy	1447	GTGTGCTCGGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1506
Db	1564	GTATGCCAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1623
Qy	1507	AGCCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1566
Db	1624	AGCCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1683
Qy	1567	GGGTGCTGATCGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1626
Db	1684	GGACTCTTGTATGCGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1743
Qy	1627	ACCGCCATCCCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1686
Db	1744	ACTGCCATCCCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1803
Qy	1687	TACTGCAATGAGCTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1746
Db	1804	TACTGCAATGAGCTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1863
Qy	1747	ACGATCTATGCAATGAGGAGGAGCTGATCTGATCTGATCTGATCTGATCTGATCTGAT	1806
Db	1864	ACCATCTATGATGAGGAGGAGGAGCTGATCTGATCTGATCTGATCTGATCTGATCTGAT	1923
Qy	1807	GAGCCACAGAGCTCTCCGAGGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT	1866
Db	1924	GAGCCACAAATCATTTTCCGCTGAGCTGATCTGATCTGATCTGATCTGATCTGATCTGAT	1983
Qy	1867	CTGTCTTGGGCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1926

1984	Db	CTGGTCCTGGGCATTTTCTTCTTGGCCCTGGCGCTGCTGGCATACCTTGTGTCTCCGGTAC	2043
1927	QY	CGGGTCAAGTTCAGAGATAGAGGCTTGGCCCAAGCCTGTACCCACGCCCTGCAGCAGGA	1986
2044	Db	CGGGTCAAGTCTCAGAGATAGAGCCTGGCTCTACCTCTTC-----TGGGCAGGA	2093
1987	QY	AGCCCCAGTCCAGAGCCCTTTGGAGCTGTTTTAACTTATATAGACTTGGGACHTGGTTCC	2046
2094	Db	AGCCCCAGTTCAGAGCCCTTTGGGGCTGTTTTAACTTGT--ACTTTGGGCATTTGGTTGCT	2151
2047	QY	GGCGGGCTATCCTCTCTCCCTTGGCTCCTCCACAGGCTGGCTGTC--GGACTGGCT	2103
2152	Db	GGTGCAGCTGCCCTCTCTCCCTGGGCTGCTACTCAGGCTGGCTGGCTGTGGGCTGTGTC	2211
2104	QY	CCAGGCTGGGCTCTGGGAGTGGGGCTCCAGAGCCCTCCCACTATGCCCAGGAGTCTTCC	2163
2212	Db	TCCCAGTCTCAGCCCGGAGGTAGGGCTCCAGCCCTCCGTC--AAGGCCAGGAGTCTTCC	2270
2164	QY	CAAGTTGATGCGGTTGTAGCTTCTCTCCCTACTCTCTCAACACCTTGCATTCGCAAGACTA	2223
2271	Db	CAAGTTGATGCACTTGTGTAGCTTCTTCCCTACTCTCTCAACACCTGCATTCGCAAGACTG	2330
2224	QY	CTGGGAGGTGCTGCTCTCTCTCCCTGCCATGGGCACCTCCTCTGCTGTCTGCCCTGGAGC	2283
2331	Db	CTGGGAGGTGCTGCTCCCTCTCTCTCCGGAAGCACCTCCTCTGTTGTTTGGCTGGGAGC	2389
2284	QY	CCTAGGCTCTTAGGGCCCCACTTACAACCTGCACAAAGTGGCCCCCTCTGGGGTCCCCA	2343
2390	Db	TCCACGTTCTTAGGCTTTCACCTTACAAGTGATCAGATTGACTTCCCTCTGGGGTCCCG	2449
2344	QY	CCACACAAGTGTTGTAAACTGGGCTGCTATAAGTTGGAGTTCAGGGCTGGGCCCTGG	2403
2450	Db	CCACAC-----TCTGCTGCTCAGTTGGAGTCCAGGGCTGGGCCCTGG	2492
2404	QY	TGGAGTCCACTGGAAGTCCCATTTATGGATGTTGAAATGGACAGGAGGAGTCTGGAAGT	2463
2493	Db	TGGGGTCTCAGGGAGTCCCATCATGGACATTGAAATGGAGCAAGAGGGAGTCTGGAAGC	2552
2464	QY	CTCTTCTCTCCTCTCTCTCTCTCCACCCCTAGACCTTGGCTGACTTGGACAAATCTGCC	2523
2553	Db	TTCTCTCTTCTCTTCTCTCTCCCTCCAAACACC--CACTTTGGCTTAATTTGGACTATCTGTT	2610
2524	QY	AGGACAGAAGCTGGGTTTCTGTCTAGTTCACACTCCCAATCCTCGGGATTTGGAGAGC	2583
2611	Db	GGGACAGAAGCCGGGCTTCTGTCTGGCCACCACATCCCAATCCTGGAGATTGGAGAGAG	2670
2584	QY	CTGGGCTGTGGGATGCCCATCCCTCTCCCATCCTTGGTGGGGGGCAGGGCCTGGT	2643
2671	Db	CATAGGATGT-----CCCGGCCCTCTCCCTAAGCCTTTCGGTCTGGATAGACCTCAAC	2722
2644	QY	GGCACCTGTGCAATAATGTCTGTGTTCTCTCCACCTGCCACTGGAACCTGGAGAATGCA	2703
2723	Db	ATCACCTCTGCAATAATGTCTGTCTACTCTCCATTCGGCCACAGGACCTTGGAGATGCA	2782
2704	QY	CTTTATTTCTGGGCGGGGTGTAGTGGGGGAAGACCAACCTCTCTTCTGCTGCCCTTA	2763
2783	Db	CTTTATTTCT---GGGGGGTGTAGCGGGGAAGCCCCCATCTCTCTCTTGTGCTCCTG	2838
2764	QY	ACGCATCCAGGGTCTCGTGATGTCTCTCTCCCTCTCCGGAGTGACAGGCACATACATGAGA	2823
2839	Db	GGGCTCAGTTTCCACAGGCTCTTTTCTCTCCAGAAATCACATCAGGCAACAAAGCCG	2898
2824	QY	ACAGGCATCTCAGCCCTACACACTTGCCATCCCTACAGCACAGAGGAGAGTGTGTT	2883
2899	Db	CCAGGCATCTCAGCCCTACACAGGCACCATCCCTTGTGGCTCAGAGGAAGAGTGGCCGT	2958
2884	QY	GGCATGCTGGTGTGGCGGTGCTGTGGTGGGAGACAGTGCACACCTCTCTCTGGGATCC	2943
2959	Db	GGCA-----GGGAGCAGTCCCAAGCTCTCTCTCCAGATC	2993
2944	QY	CATGTTGGAGACTCTAAGGATAGGCTGTGTCTGCCAGGGTGTCTACAGGAACATCGAGG	3003
2994	Db	CCACTTGT--GAGGCTAAGATGAGCTGTGACGCCAGTATGACCTATGACCTCCAGGCTCGG	3052

Qy	3004	TGTC	TACCC	CCCAAGTCTTCCCTCCTCCCAAGCAGGGTGGCAGAGGCATAGATCCCT	3066
Db	3053	TGTC	TATCT	CCCCAGTCTTCCCTCCCTTTCAGGCCCGGG-----GTGGCGCTGGCCTTCT	3106
Qy	3064	GGAGT	T	CAGAACCAACACACAAAGCAGACACACAGGCGCATAGTTGGCCTTGGCCACTGCCAC	3123
Db	3107	GGAGG	T	CAGAACCAACCA--ACAAGCAGCAAGAATGGGCATAGCTGGCCTCGGCCACTGCTCCC	3164
Qy	3124	CCAGG	CCCTCCTTTT	TGTGCTCCATGTGCGCATCTTCACTTCGCCCTACCCCTTCCCGAGC-	3182
Db	3165	CCATG	CGCTCTTT	TGTGCCCCAAGCTGACA--CTCACTCACTGCGCCTCTCCAGCT	3222
Qy	3183	CAC	TGCTGCTCATTTCAAACTTCT	TGTCATGTCCTCCCTCCACTGTCCTATCAGCAGGTGGCC	3242
Db	3223	GCCTG	CTGCTCACTCAAACTTCT	TGTCATGTCCTCCCTCCACTGTCCTATCAGCAGGTGGT	3281
Qy	3243	CCTGG	GCATCAGAACAGCCTGC	CTGGCAGCAGGTGGCAGACACACTCAGACCATCTCT	3302
Db	3282	CCTGG	ACATCAGAGCGGGGGCTG	GACATCAGGTGGCAGACAC-----CTGTC	3333
Qy	3303	GGCTT	TCTGCTGGTCCAGGCTCATTTCT	GCTTCTGATTTTCCCTCCCGCAGGGCTCAVT	3362
Db	3334	TATCT	TCTTGTGGTCCAGGATCATCT	TATCTGCTTCTCTCTCCAGGACTCTTT	3393
Qy	3363	TTCCCC	CTTTTCTGTACATCCCTG	CTACTGCTCTCACCTGCCACAGTCTTC	3422
Db	3394	CCC-----	-----CAGTCCCTGCTCACTCCT	CTGTATCTGCCCGAGTTCTTC	3438
Qy	3423	CTATC	CACAGGATGCAGTGTG	TATTTGGG	3455
Db	3439	TCATT	CACAGAGA--CCAGT	TGATTTGGG	3469
RESULT 6					
MMU426047					
LOCUS					
DEFINITION	MMU426047	3520 bp	mRNA	linear	ROD 05-FEB-2002
ACCESSION	AJ426047				
VERSION	AJ426047.2	GI:18496084			
KEYWORDS	ABC family transporter; ABCG4 gene;				white family ABC-transporter (ABCG4 gene).
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE					
AUTHORS					Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE					Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL					
REFERENCE					
AUTHORS					Oldfield, S., Lowry, C.A. and Lightman, S.L.
TITLE					Cloning and expression of a mammalian white family
JOURNAL					ABC-transporter, ABCG4
REMARK					Unpublished
REFERENCE					
AUTHORS					Oldfield, S.
TITLE					Direct Submission
JOURNAL					Submitted (19-DEC-2001) Oldfield S., URC for Neuroendocrinology,
REMARK					University of Bristol, Bristol Royal Infirmary, Marlborough Street,
REFERENCE					Bristol, BS2 8HW, UNITED KINGDOM
AUTHORS					revised by [3]
TITLE					3 (bases 1 to 3520)
JOURNAL					Direct Submission
REMARK					Submitted (30-JAN-2002) Oldfield S., URC for Neuroendocrinology,
TITLE					University of Bristol, Bristol Royal Infirmary, Marlborough Street,
JOURNAL					Bristol, BS2 8HW, UNITED KINGDOM
COMMENT					On Feb 4, 2002 this sequence version replaced gi:17976861.
FEATURES					
source					Location/Qualifiers
					1. .3520
					/organism="Mus musculus"
					/db_xref="taxon:10090"
					/clone IMAGE:4511456"
					/tissue_type="retina"
					123. .2063
					/gene="ABCG4"
gene					

CDS	123.. 2063	QY	727	TCCTCATGAAGTCCCTGGCACAGGGGGCCGTTACCATCATCTGCACCATCCACGCCC	786
/gene="ABC4"	/codon_start=1	Db	843	TCCTCATGAAGTCCCTGGCCCATGGGGCCGACCGTCACTGCACCATCCACGCCC	902
/product="putative white family ABC-transporter"	/protein_id="CAD19779.2"	QY	787	AGTGCCAAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTTGAGCCAGGGTCAAGTCAATC	846
/db_xref="GI:18496085"		Db	903	AGTGCCAAAGCTCTTTGAGATGTTTGACAAGCTCTACATCTTGAGCCAGGGGCAATGCATC	962
		QY	847	TTCAAAGGAGTGTGTACCAACCTGATCCCTATCTAAAGGGACTCCGGCTTGCAATGCCCCC	906
		Db	963	TTCAAAGGAGTGTGTACCAACCTGATCCCTATCTAAAGGGGGTGGCTTGCAATGCCCCC	1022
		QY	907	ACCTACCAACACCCGGCTGACTTTCATCATCGAGGTGGCTCTGGCCAGTATGAGACCTG	966
		Db	1023	ACCTACCAACACCCGGCTGACTTTCATCATCGAGGTGGCTCTGGCCAGTATGAGACCTG	1082
		QY	967	AACTCCATGTTGTTAGGGCTGTGCAGAAATGGGCTGTGGCTATAGGCTGAGAAGAAGC	1026
		Db	1083	AACTCCATGTTGTTAGGGCTGTGCAGAAATGGGCTGTGGCTATAGGCTGAGAAGAAGC	1142
		QY	1027	AGCCCTGAGAAGAACGAGGTCCCTGCCCATGCCCTCTGTCTCTCCGGAAGTGGATCCC	1086
		Db	1143	AGCCCTGAGAAGAACGAGGTCCCTGCCCATGCCCTCTGTCTCTCCGGAAGTGGATCCC	1202
		QY	1087	ATTGAAAGCCACACCTTTGCCACGACCCCTCACACAGTTCTGCATCTCTTCAAGAGG	1146
		Db	1203	ATTGAAAGCCACACATTTGCCACGACCCCTAAACCCAGTTCTGCATCTCTTCAAGAGG	1262
		QY	1147	ACCTTCTGTCATCTCAGGACACGGTCTCTGACCCACCTACGGTTCATGTCCCAGGTG	1206
		Db	1263	ACCTTCTGTCATCTCAGGACACGGTCTCTGACCCACCTACGGTTCATGTCCCAGGTG	1322
		QY	1207	GTATTGGGCTGCTCATCGGCTCTCTACCTGCATATTGGGACGATGCGCAAGGTC	1266
		Db	1323	GTATTGGGCTGCTCATCGGCTCTCTACCTGCATATTGGGACGATGCGCAAGGTC	1382
		QY	1267	TTCAACAACACCGGCTGCTCTTCTCTCATGCTCTCTCTCATGTTTCGCGCCCTCATG	1326
		Db	1383	TTCAACAACACCGGCTGCTCTTCTCTCATGCTCTCTCTCATGTTTCGCGCCCTCATG	1442
		QY	1327	CCAACTGTGCTACCTTCCCTTTAGAGATGGCGGTCTTTCATGAGGAGGACCTCAACTAC	1386
		Db	1443	CCAACTGTGCTACCTTCCCTTTAGAGATGGCGGTCTTTCATGAGGAGGACCTCAACTAC	1502
		QY	1387	TGTACAGCTCAAAAGCTATTACCTGGCCAAAGCATGCTGCTGCTGCTTTCAGGTG	1446
		Db	1503	TGTATACCTCAAAAGCTATTACCTGGCCAAAGCATGCTGCTGCTGCTTTCAGGTG	1562
		QY	1447	GTGTGTCGGTGTCTTACTGACGATGTTGTACTGGATGACGGGCGAGCCGCTGAGACC	1506
		Db	1563	GTATGCCAGTGTGTACTGACGATGTTGTACTGGATGACGGGCGAGCCGAGACC	1622
		QY	1507	AGCGGCTTCTGCTCTTCTCAGCCCTGCGCCACCGCCCTTGGTGGCCCAATCTTGG	1566
		Db	1623	AGCGGCTTCTGCTCTTCTCAGCCCTGCGCCACCGCCCTTGGTGGCCCAATCTTGG	1682
		QY	1567	GGGCTGCTGCTGAGGAGTCTTCCCACTCCCTACAGGTGGCCACCTTTTGGGCCCACTT	1626
		Db	1683	GGGCTGCTGCTGAGGAGTCTTCCCACTCCCTACAGGTGGCCACCTTTTGGGCCCACTT	1742
		QY	1627	ACCGCATCCCTGCTCTTGTCTGCGGCTTCTTGTGTCAGGTTCAAGACCACTCCCACT	1686
		Db	1743	ACTGCCATCCCTGCTCTTGTCTGCGGCTTCTTGTGTCAGGTTCAAGACCACTCCCACT	1802
		QY	1687	TACCTGCAATGAGGCTCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1746
		Db	1803	TACTTGAATGAGGCTCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1862
		QY	1747	AGCATATGAGGCTGAGGAGGAGACCTGACATGTTTATAGGAAGACCTGCTGCTGCTG	1806
		Db	1863	AGCATATGAGGCTGAGGAGGAGACCTGACATGTTTATAGGAAGACCTGCTGCTGCTG	1922
		QY	1807	GAGCCACAGAGCATCTCCGAGCGGTGGATGTGGAGGATGCCAAGCTCTACATGGACTTC	1866
BASE COUNT	675 a 1078 c 917 g 850 t				
ORIGIN					
Query Match	61.5%; Score 2125; DB 10; Length 3520;				
Best Local Similarity	80.4%; Pred. No. 0;				
Matches 2776; Conservative	0; Mismatches 565; Indels 112; Gaps 20;				
QY	7	ATGGCGAGAGCGGTGGAGCGGTGGGCTGTGGACTAGGGCGGGGGCTGTGGCCATG	66		
Db	123	ATGGCGAGAGGCAATTGGAGCGGTGGGCTGTGGACTAGGGCGGTGGGCGGTGGCCATG	182		
QY	67	GCGGTGACGCTGAGGACGGGGGGAACCCCTGTGCTGACACGACCTGAGAGAGGTG	126		
Db	183	GCTGTGACACTGAGGACGGGGGAGAGCCCTGTGCTGACACGACCTGAGAGAGGTG	242		
QY	127	GAGAACCATATCAGAGACAGAGCGCTTCTCTACCTGCCAAGGCGCTGCTGTGGAC	302		
Db	243	GAGAACCATATCAGAGACAGAGCGCTTCTCTACCTGCCAAGGCGCTGCTGTGGAC	362		
QY	187	ATGCAATGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	246		
Db	303	ATGCAATGCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	362		
QY	247	TATAAGACCTTCTCAAGTGCCTCTCAGTAAATCTGCGCGGGAGCTGATGGCATC	306		
Db	363	TATAAGACCTTCTCAAGTGCCTCTCAGTAAATCTGCGCGGGAGCTGATGGCATC	422		
QY	307	ATGGGCGCTCAGGCGCTGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG	366		
Db	423	ATGGGCGCTCAGGCGCTGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG	482		
QY	367	TCTGGAATGAGGGGAGATCTCTGTTTAAATGGAAGGCCACGGGAGCTGAGGACCTTCC	426		
Db	483	TCTGGAATGAGGGGAGATCTCTGTTTAAATGGAAGGCCACGGGAGCTGAGGACCTTCC	542		
QY	427	AAGATGCTCTGCTACATCATGCAAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTG	486		
Db	543	AAGATGCTCTGCTACATCATGCAAGATGACATGCTGCTGCTGCTGCTGCTGCTGCTG	602		
QY	487	GCCATGATGCTCTGCTTAACTGAACTTACTGAGAATCCCGATGTAAGAACGATCTC	546		
Db	603	GCCATGATGCTCTGCTTAACTGAACTTACTGAGAATCCCGATGTAAGAACGATCTC	662		
QY	547	GTGACAGATCTGACGGCAGTGGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	606		
Db	663	GTGACAGATCTGACGGCAGTGGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	722		
QY	607	CTCTCTGGGGGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	666		
Db	723	CTCTCTGGGGGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	782		
QY	667	GTGATGCTTCTGATGAGCCACAGTGGTCTGATGAGGCTGCTGCTGCTGCTGCTGCTG	726		
Db	783	GTGATGCTTCTGATGAGCCACAGTGGTCTGATGAGGCTGCTGCTGCTGCTGCTGCTG	842		

2957	GGCA-----GGGAGCAGTCCCAACGTCCTCTCCGCAATC	2999
2944	CATGTTGGAGACTCTAAGATTAAGCTGGTGTGCCAGGGTGCTCTACAGAACTCGCAGG	3003
2992	CNACTTGG-GAGGCTAAGATGAGACTGGTGTGACCCAGTATGACTCCAGGCTTCGGCAGG	3050
3004	TGCTTACCCCAAGTCTTCCCCTCCCAAGCCAGGGGTGGCAGGGCACTAGATCCCT	3063
3051	TGTCTATCTCCAGTCTCCCTCCCTCTTGAGCCCGGG-----GTGGCGCTGGCCCTCT	3104
3064	GGAGTTTCAGGAACCAACAGCAAGCAACACGGGCATAGTTGGCTTGGCCACTGGCCAC	3123
3105	GGAGGTTCAGAACCCA--ACAAGCACAGAATGGGCATAGCTGGCCCTCGGCCACTGTCTCC	3162
3124	CCAGGCGCTCCTTTTGTGCTCCATCGTGCATCTTCACCTCCCTACCCCTTCCCCAGC-	3182
3163	CCATGGCCCTCCTTTTGTGCCCCAGCTGACA--CTCACTTCACTCGCCCTCTCCAGCT	3220
3183	CAGTGTGCTCATTTCAAACCTTCTGTCCATGTCCCTCCACTGTTTCTTATCAGCAGGTGGCC	3242
3221	GCCTGCTGCTCACTCAAACTTCTGTTTCACATGATTTCA-TCCTTCGGTCAGCAGCTGGTT	3279
3243	CTGGGCATCAGAACAGCTTGCCTTGGCCACCAAGTGGCAGACACTCAGACGATGTCT	3302
3280	CCTGGACATCAGAGCGCGGGCTGGACATCAGGTGGCAGACACAC-----CTGTCT	3331
3303	GGCTTTCCCTGGTGGGTCCAGGCTCATTTCTGATTTCCCTCCCCCAGGGCTCATTT	3362
3332	TATCTTTCTTGGTGGGTCCAGGATCATCCCTATATCTGTGTTCTCCTCTCCAGGACTCTTT	3391
3363	TTCCCCCTTTTCTCTGTACACATCCCTGTCTAGCTCTCTCACCTGCCACAGATTCTTC	3422
3392	CCC-----CAGTCCCTGTGCTCACTCCTCTGTATCTGCCCCAGGTTCTTC	3436
3423	CTATCACACAGGATGCCAGTTGATTTGGG	3455
3437	TCATTACACAGAGA--CCAGTTCTATTTTGGG	3467

RESULT 7
 AP001315
 LOCUS
 DEFINITION
 AP001315
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

AP001315
 Homo sapiens chromosome 11 clone CMB9-93G21 map 11q23, WORKING
 DRAFT SEQUENCE, 56 unordered pieces.
 AP001315
 GI:11071931
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 Homo sapiens DNA, clone:CMB9-93G21.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 157405)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Homo sapiens 157,405 genomic DNA of 11q23
 Published Only in DataBase (2000)
 2 (bases 1 to 157405)
 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
 Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (01-MAR-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 On Nov 1, 2000 this sequence version replaced gi:1117242.
 ----- Genome Center
 Center: RIKEN Genomic Sciences Center(GSC)
 Center code: RIKEN
 Web site: http://hgp.gsc.riken.go.jp/
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11

Center clone name: CMB9-93G21
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 130262 bases at least Q40
Consensus quality: 136359 bases at least Q30
Consensus quality: 143933 bases at least Q20
Insert size: 151905; sum-of-contigs
Quality coverage: 8.32x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 56 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1
7883 contig of 7883 bp in length
7884 14891 contig of 6908 bp in length
14892 22645 contig of 7654 bp in length
22746 29670 contig of 6925 bp in length
29771 36703 contig of 6933 bp in length
36804 43057 contig of 6254 bp in length
43158 48313 contig of 5156 bp in length
48414 53404 contig of 4991 bp in length
53505 57461 contig of 3957 bp in length
57562 61460 contig of 3899 bp in length
61561 64861 contig of 3086 bp in length
64962 71295 contig of 3148 bp in length
71396 74583 contig of 3188 bp in length
74684 77136 contig of 2453 bp in length
77237 81080 contig of 3844 bp in length
81181 84386 contig of 3206 bp in length
84487 87351 contig of 2865 bp in length
87452 88841 contig of 1390 bp in length
88942 89941 contig of 100 bp in length
89954 90953 contig of 1912 bp in length
90954 93372 contig of 2419 bp in length
93373 93472 contig of 100 bp in length
93473 95647 contig of 2175 bp in length
95648 95747 contig of 100 bp in length
95748 97782 contig of 2035 bp in length
97783 97882 contig of 100 bp in length
97883 100347 contig of 2465 bp in length
100348 100447 contig of 100 bp in length
100448 102232 contig of 1785 bp in length
102233 102332 contig of 100 bp in length
102333 105032 contig of 2700 bp in length
105033 105132 contig of 100 bp in length
105133 108433 contig of 3301 bp in length
108434 108533 contig of 100 bp in length
108534 110309 contig of 1776 bp in length
110310 110409 contig of 100 bp in length
110410 112684 contig of 2275 bp in length
112685 112785 contig of 100 bp in length
112785 114376 contig of 1592 bp in length
114377 114476 contig of 100 bp in length
114477 116547 contig of 2071 bp in length
116548 116647 contig of 100 bp in length
116648 118469 contig of 1822 bp in length
118470 118569 contig of 1048 bp in length
118570 118669 contig of 1048 bp in length

155908 157405 contig of 1498 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
7883 7883 contig of 7883 bp in length
7884 7983 gap of 100 bp
14891 14891 contig of 6908 bp in length
14892 14991 gap of 100 bp
22645 22645 contig of 7654 bp in length
22746 22745 gap of 100 bp
29670 29670 contig of 6925 bp in length
29771 29770 gap of 100 bp
36704 36703 contig of 6933 bp in length
36804 36803 gap of 100 bp
43058 43057 contig of 6254 bp in length
43158 43157 gap of 100 bp
48314 48313 contig of 5156 bp in length
48414 48413 gap of 100 bp
53405 53404 contig of 4991 bp in length
53505 53504 gap of 100 bp
57461 57461 contig of 3957 bp in length
57562 57561 gap of 100 bp
61460 61460 contig of 3899 bp in length
61561 61560 gap of 100 bp
64862 64861 contig of 3301 bp in length
64962 64961 gap of 100 bp
68048 68047 contig of 3086 bp in length
68148 68147 gap of 100 bp
71295 71295 contig of 3148 bp in length
71396 71395 gap of 100 bp
74584 74583 contig of 3188 bp in length
74684 74683 gap of 100 bp
77137 77136 contig of 2453 bp in length
77237 77236 gap of 100 bp
81080 81080 contig of 3844 bp in length
81181 81180 gap of 100 bp
84387 84386 contig of 3206 bp in length
84487 84486 gap of 100 bp
87352 87351 contig of 2865 bp in length
87452 87451 gap of 100 bp
88841 88841 contig of 1390 bp in length
88942 88941 gap of 100 bp
90954 90953 contig of 1912 bp in length
90954 93372 contig of 2419 bp in length
93373 93472 gap of 100 bp
93473 95647 contig of 2175 bp in length
95648 95747 gap of 100 bp
95748 97782 contig of 2035 bp in length
97783 97882 gap of 100 bp
97883 100347 contig of 2465 bp in length
100348 100447 gap of 100 bp
100448 102232 contig of 1785 bp in length
102233 102332 gap of 100 bp
102333 105032 contig of 2700 bp in length
105033 105132 gap of 100 bp
105133 108433 contig of 3301 bp in length
108434 108533 gap of 100 bp
108534 110309 contig of 1776 bp in length
110310 110409 gap of 100 bp
110410 112684 contig of 2275 bp in length
112685 112784 gap of 100 bp
112785 114376 contig of 1592 bp in length
114377 114476 gap of 100 bp
114477 116547 contig of 2071 bp in length
116548 116647 gap of 100 bp
116648 118469 contig of 1822 bp in length
118470 118569 gap of 100 bp

Db 1567 GCAGACACACTCAGACGATGTCGGCTTCCTCGTGGGTCCAGGCTCACTTCTGCTTCTGA 1622

QY 3340 TTTTCCCTCCCCCAGGGCTCATTTTCCCTTTTCCCTTTTCTGTACACATCCCTGTCTACTCTCC 3399
 |||||
 Db 1627 TTTTCCCTCCCCCAGGGCTCATTTTCCCTTTTCTGTACACATCCCTGTCTACTCTCC 1686
 |||||

QY 3400 TCTACCCCTGCCACAGATTCTTCCCTATCACACAGGATGCCAGTGTGATTGTTGGG 3455
 |||||
 Db 1687 TCTACCCCTGCCACAGATTCTTCCCTATCACACAGGATGCCAGTGTGATTGTTGGG 1742
 |||||

RESULT 9
 AP002956 219574 bp DNA linear PRI 15-AUG-2002
 LOCUS Homo sapiens genomic DNA, chromosome 11q clone:R1105h09, complete
 DEFINITION sequence.
 ACCESSION AP002956
 VERSION 1 GI:22255355
 KEYWORDS HTG.
 SOURCE Homo sapiens DNA, clone:R1105h09.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Homo sapiens 219,574 genomic DNA of 11q
 JOURNAL Published Only in Database (2002),
 REFERENCE 2 (bases 1 to 219574)
 AUTHORS Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (24-NOV-2000) Masahira Hattori, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 COMMENT Overlapping Clones: pD1A4, PQ38G8
 Cytogenetic Position: 21q22.1, region:D21S226-AML
 STS Markers (ePCR): stSG53747, SHGC-16045, D11S2450, SHGC-7143,
 WI-16056,
 stSG31054, sts-X57110, WI-8978, GDB:198117, stSG29284, SHGC-130657,
 D11S1941E, WI-14589, stSG60191, stSG50535, A002048, stSG26946,
 stSG41092,
 TIGR-A005D38, Bdaa7g02, D11S869E, RH80030
 Contamination: none detected
 non-AGCT bases: none
 Additional author information
 Arai,Y., Kubo,T., Ohki,M.
 National Cancer Center Research Institute
 Cancer Genomic Division
 5-1-1, Tsukiji, Chuo-ku, Tokyo, JAPAN
 zip: 104-0045
 phone: +81-3-3542-2511 ex4752, fax: +81-3-3542-0688 e-mail:
 yara@ncrc.go.jp.
 FEATURES
 source Location/Qualifiers
 1. 219574
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11q"
 /clone="R1105h09"
 BASE COUNT 54883 a 50356 c 53844 g 60491 t
 ORIGIN

Query Match 50.1%; Score 1729.4; DB 9; Length 219574;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1753; Conservative 0; Mismatches 21; Indels 1; Gaps 1;

QY 1681 CCACCTTACCTGCAATGGAGCTCCTTCTCTCTCTATGTCAGGTATCGCTTTGAGGGTG 1740
 |||||
 Db 30111 CCTGTGTGTGATGCCATCTGATATCCCTGCCTGCCTAGGTATGCTTTGAGGGTG 3017

Db	31250	AGAACAGGCATCTCAGCCCTACACATTGCCATCCCTACAGCACAGGAAGTGTAT	31309
Qy	2881	GFTGGCATCTGTTGGTGGCGGGTGCTGTTGGAGGACAGTGCCAACTCTCTCTGGGGA	2940
Db	31310	GFTGGCATCTGTTGGTGGCGGGTGCTGTTGGAGGACAGTGCCAACTCTCTCTGGGGA	31369
Qy	2941	TCCCATGTTGGAGACTCTAAGGATPAAGCTGTGTGCTGCCAGGTGTCTACAGAACTGC	3000
Db	31370	TCCCATGTTGGAGACTCTAAGGATPAAGCTGTGTGCTGCCAGGTGTCTACAGAACTGC	31429
Qy	3001	AGGTGTCTACCCCCCAAGTCTTCCCTCCTCCAAAGCCAGGGTGGCACAGGCACTAGATC	3060
Db	31430	AGGTGTCTACCCCCCAAGTCTTCCCTCCTCCAAAGCCAGGGTGGCACAGGCACTAGATC	31489
Qy	3061	CCTGGAGTTACGAAACCAACACAGCACACACAGGGGATPAAGTTGGCTTTGGCCACTGC	3120
Db	31490	CCTGGAGTTACGAAACCAACACAGCACACACAGGGGATPAAGTTGGCTTTGGCCACTGC	31549
Qy	3121	CACCCACGGCCCTCCTTTTGTGCTCCATGCTGGCATCTTCACTCCCTTACCCCTTCCCA	3180
Db	31550	CACCCACGGCCCTCCTTTTGTGCTCCATGCTGGCATCTTCACTCCCTTACCCCTTCCCA	31609
Qy	3181	GCACGTGCTGCTATTCAAACTTCTGTCCATGTCCCTCCACTGTTCCTATCAGCAGTGG	3240
Db	31610	GCACGTGCTGCTATTCAAACTTCTGTCCATGTCCCTCCACTGTTCCTATCAGCAGTGG	31669
Qy	3241	CCCTGGGCATCAGAACACGCTGCCCTGGGCACAGGTGGCAGACACACTCAGAGCATGT	3300
Db	31670	CCCTGGGCATCAGAACACGCTGCCCTGGGCACAGGTGGCAGACACACTCAGAGCATGT	31729
Qy	3301	CTGGCTTTCTGGTGGGTCCAGGCTCATTTCTGTCTGATTTCCCTCCCCAGGGCTCA	3360
Db	31730	CTGGCTTTCTGGTGGGTCCAGGCTCATTTCTGTCTGATTTCCCTCCCCAGGGCTCA	31789
Qy	3361	TTTTCCCCCTTTTCTGTFACACATCCCTGTCTACCTCCTCTCACCCGTCACAGATTCT	3420
Db	31790	TTTTCCCCCTTTTCTGTFACACATCCCTGTCTACCTCCTCTCACCCGTCACAGATTCT	31849
Qy	3421	TCTATPACACAGGATGCCAGTTGTATTGTGGG	3455
Db	31850	TCTATPACACAGGATGCCAGTTGTATTGTGGG	31884

RESULT 10

AP000833/C

LOCUS

DEFINITION Homo sapiens chromosome 11 clone RP11-849I24 map 11q23, WORKING DRAFT SEQUENCE, 43 unordered pieces.

ACCESSION

CONCLUSION

KEYWORDS

SOURCE
ORGANISM

ORGANISM

REFERENCE 1 (bases 1 to 190762)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE AUTHORS

QUESTIONS

TITLE

JOURNAL

REFERENCE

AUTHORS

T.T.T.F.

JOURNAL
TITLE
Direct Submission
Submitted (03-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN) Genomic Sciences Center (GSC).

and Chemical Research (KIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail: hattori@rsc.kit.ac.jp)

```

Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information -----
Center project name: HumDrafl1
Center clone name: RP11-849I24
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Er-Amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161205 bases at least Q40
Consensus quality: 174612 bases at least Q30
Consensus quality: 182186 bases at least Q20
Insert size: 186562; sum-of-contigs
Quality coverage: 4.07x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 43 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1	18426	contig of	18426	bp in length
18527	34557	contig of	18527	bp in length
34658	43481	contig of	34658	bp in length
43582	51726	contig of	43582	bp in length
51827	58795	contig of	51827	bp in length
58896	67937	contig of	58896	bp in length
68038	75037	contig of	68038	bp in length
75138	82043	contig of	75138	bp in length
82144	88404	contig of	82144	bp in length
88505	95104	contig of	88505	bp in length
95205	100820	contig of	95205	bp in length
100921	105464	contig of	100921	bp in length
105565	110699	contig of	105565	bp in length
110800	115670	contig of	110800	bp in length
115771	120227	contig of	115771	bp in length
120328	123755	contig of	120328	bp in length
123856	127538	contig of	123856	bp in length
127639	131293	contig of	127639	bp in length
131394	135587	contig of	131394	bp in length
135688	139626	contig of	135688	bp in length
139727	142738	contig of	139727	bp in length
142839	146711	contig of	142839	bp in length
146812	150501	contig of	146812	bp in length
150602	153871	contig of	150602	bp in length
153972	158170	contig of	153972	bp in length
158271	161303	contig of	158271	bp in length
161404	163262	contig of	161404	bp in length
163363	165784	contig of	163363	bp in length
165885	167904	contig of	165885	bp in length
168005	171033	contig of	168005	bp in length
171134	172663	contig of	171134	bp in length
172764	174696	contig of	172764	bp in length
174797	176092	contig of	174797	bp in length
176193	178229	contig of	176193	bp in length
178330	180717	contig of	178330	bp in length
180818	182655	contig of	180818	bp in length
182756	183975	contig of	182756	bp in length
184076	185539	contig of	184076	bp in length
185640	186997	contig of	185640	bp in length
187098	188232	contig of	187098	bp in length
188433	189332	contig of	188433	bp in length
189543	190547	contig of	189543	bp in length
190648	190762	contig of	190648	bp in length

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft'

* consists of 43 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown


```
Db 27117 GCCCCTGGGCATCAGAACAGCCTGCCCTGGCACCAGGTGGCAGACACACTCAGAGCATG 27058
QY 3300 TCTGGCTTCCCTGGTGGGTCAGGCTCATCTGCTTCTGATTTCCCTCCCGAGGCTC 3359
Db 27057 TCTGGCTTCCCTGGTGGGTCAGGCTCATCTGCTTCTGATTTCCCTCCCGAGGCTC 26998
QY 3360 ATTTCCTCCCTTTTCTGTACACATCCCTGTCTACTCTCTACCTCTCTACCTGCCACAGATTC 3419
Db 26997 ATTTCCTCCCTTTTCTGTACACATCCCTGTCTACTCTCTACCTGCCACAGATTC 26938
QY 3420 TTCCTATACACAGGATGCCAGTTGATTTGFGG 3455
Db 26937 TTCCTATACACAGGATGCCAGTTGATTTGFGG 26902

RESULT 13
AY040865
LOCUS
DEFINITION
Mus musculus ATP-binding cassette transporter ABCG4 (Abcg4) mRNA,
complete cds.
ACCESSION
AY040865
VERSION
AY040865.1 GI:19849454
WORDS
Mus musculus.
SOURCE
Mus musculus
ORGANISM
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2056)
Yoshikawa, M., Yabuuchi, H., Kuroiwa, A., Ikegami, Y., Sai, Y.,
Tamai, I., Tsuji, A., Matsuda, Y., Yoshida, H. and Ishikawa, T.
Molecular and cytogenetic characterization of the mouse ATP-binding
cassette transporter Abcg4
Gene 293 (1-2), 67-75 (2002)
JOURNAL
MEDLINE
2213307
PUBMED
12137944
2 (bases 1 to 2056)
Yoshikawa, M., Yabuuchi, H., Ikegami, Y. and Ishikawa, T.
Direct Submission
Submitted (18-JUN-2001) Biomolecular Engineering, Tokyo Institute
of Technology, 4259 Nagatsuta, Midori-ku, Yokohama, Kanagawa
226-8501, Japan
FEATURES
Location/Qualifiers
1..2056
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/sex="male"
/tissue_type="brain"
1..2056
/gene="Abcg4"
88..2028
/gene="Abcg4"
/codon_start=1
/product="ATP-binding cassette transporter ABCG4"
/protein_id="AAK91781.1"
/db_xref="GI:19849455"
/translation="MAEKALEAVGGGLGPGVAMAVTLEDGAEPPVLTTHLKKVNHII
TEAOREHLPKRSAYDIEFVELSVYREGPCWRKRGYKTLKCLSKRCRELIGIM
PSGAKSTFMNLAGYREGMKQIILVNGRPRELTFPRKMSYIQDDMLLPHLVLE
AMVSAULKLSKQVKRELVTETILALGMSCSHRTALLSGGQRKRLALELVNN-
PVMFFDSTGLDSASCFVQVSLMKSLAHGRTVICTIHQPSAKLFEMFDKLYILSQ
GQCFKPVNTLIPYKLGHLCHPTYPHNADFIIEVASGEYDGLNPLMFLRAYONGICT
MAEKSSPKNELPAHCTPCPELDPIESHTFATSLTQFCILFRRTLFSLIRDPVLT
HLRFMSHVLIGLVILGLVLIHGDDAKVFNNTGFLFSLMFLFAALMPTVLTFFLEM
AVFMEHLNRYWTLKAYLAKTMADVPQVQVPCVYVYCSIVYHMTQPAETSRFLFSA
LAIALTAQGLGLIGASLSQVATFPGVPTAIPVLLFSGFVFSFTIPYLOWNS
LYSYVRYFEGSLIYIGMERHLTCLDEQCFFRDPQIILRELDVEEAKLYMDFLVLG
IFFLALRLALLVLYRVRKSR"
398 a 611 c 560 g 487 t

BASE COUNT
ORIGIN
Query Match 47.3%; Score 1633; DB 10; Length 2056;
```


Db 1108 AGCCCTGGGAAGATGAGCTCCCTGCCACTGCCACCCTTGCCCTCCGAGCTGGATCCT 1167
QY 1087 ATTGAAGCCACACCTTTGCCACACAGCCTCACACAGTCTCTGCATCTCTTCAAGAGG 1146
Db 1168 ATTGAAGCCACACATTTGCCACACAGCCTTAACCCAGTTCTGCATCTCTTCAAGAGG 1227
QY 1147 ACCTTCTCTGCATCCTCAGGAGACAGGCTGCTGACCCACCTACGGTTCATGTCCACGCTG 1206
Db 1228 ACTTCTCTGCATCCTCAGGAGACAGGCTGCTGACCCACTTGCCTGTTTACATGTG 1287
QY 1207 GTTATGGCGTCTCATCGGCTCTCTACCTGCATATTTGGGACGATGCCAGCAGGTC 1266
Db 1288 TTGATGGCGTCTCATGCGCTCTCTACCTGCATATTTGGGACGATGCCAGCAGGTC 1347
QY 1267 TTCAACAACACCGGCT 1326
Db 1348 TTCAACAACACCGGCT 1407
QY 1327 CCAACTGTCTCACTTCCCTTTAGAGATGGCGGTCTTCATGAGGAGACCTCAACTAC 1386
Db 1408 CGGACGCTCTCACTTCCCTTTAGAGATGGCGGTCTTCATGAGGAGCAGCTTGAACAT 1467
QY 1387 TGTACAGCTCAAAAGCTATTACCTGGCCAAAGACCATGCTGACCTGCGCTTTCAGGTG 1446
Db 1468 TGTATACTCTCAAGCCCTATTACTTGGCCAAAGACCATGCTGATGCTGCTTCCAGGTG 1527
QY 1447 GTGTGCGGTGCTCTACTGACGATTTGTACTGATGACGGCCAGCGCTGAGACC 1506
Db 1528 GTATGCCAGTGGTGTACTGACGATCTGTACTGATGACAGGCCAGCCGAGACC 1587
QY 1507 AGCGGCTTCTGCTCTCTAGCCCTGGCCAGCCAGCCGCTTGTGGGCCCAATCTTGG 1566
Db 1588 AGCGGCTTCTGCTCTCTAGCCCTGGCCAGCCAGCCGCTTGTGGGCCCAATCTTGG 1647
QY 1567 GGCTGCTCATGCGAGCTCTCTCAACTCCCTACAGGTGGCCAGCTTGTGGGCCCAAT 1626
Db 1648 GGACTCTTATGGGCGCTGCT 1707
QY 1627 ACGCCATCCCTGCT 1686
Db 1708 ACTGCCATCCCTGCT 1767
QY 1687 TACTTCAATGAGCT 1746
Db 1768 TACTTCAATGAGCT 1827
QY 1747 AGCATCTATGGATGAGGAGGAGACCTGACATGTTTATAGGAAGCTGCTGCTGCTCGG 1806
Db 1828 ACATCTATGGATGAGGAGGAGACCTGACATGTTTATAGGAAGCTGCTGCTGCTCGG 1887
QY 1807 GAGCCACAGAGCATCTCCGAGCGCTGGATGTGGAGGTGCCAAGCTCTACATGGACTTC 1866
Db 1888 GACCCCAAAATCATTTTGGCTGAGCTGGATGTAGAGGAGGCCAACTCTACATGGACTTC 1947
QY 1867 CTGGCTTGGGATCTTCTCTAGCCCTCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTTAC 1926
Db 1948 CTGGCTTGGGATCTTCTCTAGCCCTCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2007
QY 1927 CGGCTCAAGTCAGAGATAGAGGCTTGGCCAGCCCTGTAC 1967
Db 2008 CGGCTCAAGTCAGAGATAGAGGCTTGGCTTCTACCTCTTC 2048

RESULT 14
LOCUS BC023077
DEFINITION Mus musculus, clone IMAGE:5359019, mRNA, linear ROD 07-AUG-2002
ACCESSION BC023077
VERSION BC023077.1 GI:18605999
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2730)
Strausberg, R.
Direct Submission
Submitted (04-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: ang@bcm.tmc.edu
Gunnarathne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 54 Row: j Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
Location/Qualifiers
1. .2730
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5359019"
/tissue_type="Eye, retina, mouse strain C57Bl/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1. .1272
/codon_start=1
/product="Unknown (protein for IMAGE:5359019)"
/protein_id="AAH23077.1"
/db_xref="GI:18606000"
/translation="DLSLDSGLDSASCFOVSLMKSLAHGRTVICTIHQPSAKLFEM
FKLYTLGSGQCIKFGVNTNLIPYLKGLGHPTVHPNADFIIEVAGGYGDLNPMIF
RQVNGCLTMAEKKSSPGKNELPAHCTPPPELDPESHFTPATSLTQFCILERTFL
SLIRDVLTLRPMNSHLVGLLGLHIGDASKVFNNTGFLFSSMLFLMFAALMP
TVITFPLENAVEMREHLNWTLYKAYLAKTMADVPVQVPCVYVYVMTGQPAE
TSRFLFSAIATAIQAQSLGLIQAATSLQVATFVGPVPAIPVLLSGFVGFPT
IPTYLQWSSYLSVVRVYFEGLLITTYGEMRHLTCLDEQCFPRDPQIILRELDEVEAK
LYMDFLVIGIFFLRLRLAYLVLRVVKSER"
BASE COUNT 514 a 858 c 660 g 698 t
ORIGIN
Query Match 45.0%; Score 1554.8; DB 10; Length 2730;
Best Local Similarity 77.9%; Pred. No. 3e-308;
Matches 2157; Conservative 0; Mismatches 502; Indels 111; Gaps 19;
QY 690 CAGTGTCTGGATAGCGCCTCTGTTTCCAGTGTGTCCCTCATGAAGTCCCTGGCACA 749
Db 15 CAGCGGTCTAGACAGCGCTTCTGTTTCCAAAGTGTGTCCCTCATGAAGTCCCTGGCCA 74
QY 750 GGGGGCCCTAGCATCATCTGCACCATCCACGAGCCAGTCCAGCTCTTTGAGATGTT 809
Db 75 TGGGGCCCGCCAGCTCATCTGCACCATCCACGAGCCAGTCCAGCTCTTTGAGATGTT 134
QY 810 TGACAAGCTCTACATCCTGAGCAGCGGTGAGTTCATCTTCAAAGAGGTGCTCACCAACCT 869
Db 135 TGACAAGCTCTACATCCTGAGCAGCGGCAATGCATCTTCAAGGGGTGCTTACCAACCT 194
QY 870 GATCCCTTATCAAAGGAGTGGCTTGGATTGCCCCACCTACCAACACCGGCTGACTT 929
Db 195 GATTCCCTATCTGAAGGGCTTGGCTTGGCACTGCCACCTACCAACACCGGCTGACTT 254

Search completed: June 16, 2003, 16:23:41
Job time : 8643 secs

Qy	919	CCGGCTGACTTTCATCATCGAGAGTGGCCCTCTGGCGAGTATGGAGACCTTGAACCCCATGTTG	978
Db	1069	CCAGCAGACTTTGTCTATGGAAGTGGCATCAGGGAGTACGGGGATCAGAAGACAGTCGCGCTG	1128
Qy	979	TTCAGGGCTGTGCAGAAATGGCTGTGCGCTATGGCTGAGAAGAGACGCCCTGAGAAG	1038
Db	1129	GTGAGACCGTGGAGAGGCATGTGTGACGCTGACTATTAAGAGACCTCGGGGGCGAC	1188
Qy	1039	AACGAGTCCCTGCCCCATCCCTTCTTCTCCGGAAGTGGATCCCATTTGAAA	1093
Db	1189	ACCGATGTAACCGTTTCTTTTGGCACCGGCTGCTGAAGAGGACTCCGCCCTCCATGGAA	1248
Qy	1094	----GCCACACCTTTGCCACACGACCCCTCACACAGTTCTTGATCCTCTTCAAGAGGACC	1149
Db	1249	GGTTGGCCATAGCTTCTCGGCCAGCTGCTCCACCACTGTTCTGCATCCTCTTCAAGAGGACC	1308
Qy	1150	TTCTCTGCCATCCTCAGGGACACGGTCTCACCACACCTACGGTTTCATGTCCCACTGGTT	1209
Db	1309	TTCTCAGCATCATCGGGACTCGGTCTTGACACATCTGCGAATCACTCGACATTTGGG	1368
Qy	1210	ATTGGGCTGCTCATCGGCGCTCTCTACCTGCATATTGGCGACGATGCCAAGGCTTTC	1269
Db	1369	ATCGGCTGTCTCATTTGGCGCTCTGTCTACCTGGGANTTGGGAATTAAGCCAAAGAAGTCCTT	1428
Qy	1270	AACAACACCGGCTGCTTCTTCTCCATGCTCTCTCTCATGTTCCGCGCCCTCATGCCCCA	1329
Db	1429	AGCACTCCGGCTCTCTGTTCTCATGCTGTCTCTCATGTTGCTGCCCTCATGCCCC	1488
Qy	1330	ACTGTGCTACCTTTCCCTTTAGAGATGGCGTCTTCTATGAGGAGCACCTTCAACTACTGG	1389
Db	1489	ACTGTTCTGACCTTTCCCTTGGAGATGAGTGTCTTCTCTCGGGAGCACCTTGAACACTACTGG	1548
Qy	1390	TACAGCTCAAAAGGATTAACCTTGGCCAAAGACCATGGCTGACGTGCGCTTTTCAGTGGTG	1449
Db	1549	TACAGCTCAAGGCGCTACTACCTTGGCAAGACCATGGCCGATGTCCCTTTTCAGATCATG	1508
Qy	1450	TGTCGGTGTGCTACTGACGATTTGTCTACTGTGATGACGGCCACCGCTGAGACCAAGC	1509
Db	1609	TTCCCTGTGSCCTACTGCGATATCGTATACTGGATGACGTCCCGACGCTGGAGCGTGTG	1568
Qy	1510	CGCTTCCTGCTCTTCTCAGCCCTGGCCACCGCCACCGCCTTGTGGGCCAATCTTTGGGG	1569
Db	1669	CGTTTTGTGCTGTTCGCTGCTTGGGTACCATGACATCGCTGGTGGCCAGTCTCTTAGGA	1728
Qy	1570	CTGCTGATCGAGCTGCTTCCAACTCCCTACAGTGGCCACTTTTGTGGGCCAGTTACC	1629
Db	1729	CTACTGATGGAGCTGCATCCACATCCCTGCAGGTTGCGACATTTGTGGTCCCGTGACA	1788
Qy	1630	GCCATCCCTGTCTCTTGTTCGCGCTCTTTGTTCAGCTTCAAGACCATTCCCCACTTAC	1689
Db	1789	GCCATCCCGTCTGCTCTTCTCCGANTCTTTGTGAGCTTTGACACCATCCCGACCTAC	1848
Qy	1690	CTSCAATGAGCTCTATCTCTCTATGTCAGGTATGGCTTTCAGGGTGTGATCCTTCAGC	1749
Db	1849	CTCAGTGGATGTCTTACATCTCCTATGTCAGATACGGCTTTGAGGGGTTCATCTTGTC	1908
Qy	1750	ATCTATGGCATGGAGGAGGAGACCTTGACATG---TTTAGAGGAACGCTGCCGTTCCCG	1806
Db	1909	ATCTACGGCTTGGACCGGAAGACCTGCACTGGACATCGCGAGACATGCCACTTCCAG	1968
Qy	1807	GAGCCACAGACATCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATGGACTTTC	1866
Db	1969	AAGTCAGAGGCCATCTCAGGGAGCTTGGAGCTGGAGAATCCGAAGCTGACCTTGGATTTC	2028
Qy	1867	CTGGTCTTGGGATCTTCTTCTTAGCCCTCGGCGTCTGGCCTACCTTGTGCTGCCCTTAC	1926
Db	2029	ATCGTCTTGGGATCTTCTTCTATCTCCTTGGCGCTCATCGCCTATTTGTGCTTCAGATAC	2088
Qy	1927	CGGGTCAAGTCAAGAGATAGAGGCTTGGCCCCAGCGCTGTACCCCCAGCCCTGCAG	1981
Db	2089	AAATCCGGGCTGAGAGTTAAACCCGCTCTCAAGCCAGCAACGAGCAAGAC	2143

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:39:13 ; Search time 4538 Seconds
(without alignments)
12330.411 Million cell updates/sec

Title: US-10-072-621-4
Perfect score: 3455
Sequence: 1 gccaccatgcccgaagagc.....atgccagttgtattgtggg 3455

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba.*

2: em_estba.*

3: em_estin.*

4: em_estin.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_htc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: gb_gss.*

18: em_gss_hum.*

19: em_gss_inv.*

20: em_gss_pln.*

21: em_gss_vrt.*

22: em_gss_fun.*

23: em_gss_mam.*

24: em_gss_mus.*

25: em_gss_other.*

26: em_gss_pro.*

27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	19.9	1856	11	BC030301 Mus muscu
2	634.8	18.4	851	14	BQ179452 UI-M-EMO-
3	603	17.5	778	13	BI731100 603351741
4	548	15.9	548	14	BQ636356 hd07h11.y
5	541.6	15.7	1045	14	BM925745 AGENCOURT
6	506.8	14.7	1428	11	AK018283 Mus muscu

7	463	13.4	820	12	BG298084
8	440.4	12.7	777	9	AU119946
9	406.6	11.8	974	12	BG292762
10	397.4	11.5	964	12	BE742274
11	383.2	11.1	885	14	BQ930068
12	376	10.9	879	12	BF159563
13	369	10.7	688	17	BQ6468
14	367.4	10.6	625	14	BQ769103
15	360.2	10.4	647	10	BE262142
16	352.4	10.2	747	12	BE315332
17	352	10.2	594	10	BE148782
18	345	10.0	810	13	BI736747
19	325	9.4	630	10	BB640447
20	313.6	9.1	695	14	BQ745197
21	313.6	9.1	720	10	BE260763
22	313	9.1	535	12	BE861999
23	305.4	8.8	716	12	BG747171
24	297.8	8.6	537	14	BQ637714
25	294.4	8.5	802	14	W21981
26	291.6	8.4	476	10	AW254434
27	291.6	8.4	482	9	AI712719
28	271.6	7.9	440	10	AW253072
29	264	7.6	426	9	AI711009
30	263.8	7.6	684	12	BG283492
31	262.6	7.6	422	9	AI712660
32	261.8	7.6	667	10	BB630594
33	258	7.5	567	9	AA944183
34	257.4	7.5	443	12	BF654958
35	255.8	7.4	554	12	BE948744
36	249.2	7.2	650	14	BQ640204
37	246.4	7.1	540	10	AW491055
38	242.8	7.0	400	9	AI716340
39	242.6	7.0	608	10	AV973769
40	242	7.0	601	12	BG083855
41	240.6	7.0	709	13	BM593317
42	238.6	6.9	371	14	F06912
43	234	6.8	304	14	BM692180
44	230.2	6.7	686	13	BM620534
45	229.6	6.6	500	14	BQ258653

ALIGNMENTS

RESULT 1
BC030301

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC030301 1856 bp mRNA linear HTC 07-AUG-2002
Mus musculus, Similar to RIKEN cDNA 6430517004 gene, clone
IMAGE:5367539, mRNA.

BC030301 GI:20987866

BC030301 HTC

BC030301 house mouse.

BC030301 Mus musculus

BC030301 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

BC030301 Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

BC030301 1 (bases 1 to 1856)

BC030301 Strausberg, R.

BC030301 Direct Submission

BC030301 Submitted (07-MAY-2002) National Institutes of Health, Mammalian

BC030301 Gene Collection (MGC), Cancer Genomics Office, National Cancer

BC030301 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

BC030301 USA

BC030301 NIH-MGC Project URL: http://mgc.nci.nih.gov

BC030301 Contact: MGC help desk

BC030301 Email: c9apbs@mail.nih.gov

BC030301 Tissue Procurement: The Cepko Laboratory

BC030301 CDNA Library Preparation: Life Technologies, Inc.

BC030301 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

BC030301 DNA Sequencing by: National Institutes of Health Intramural

BC030301 Sequencing Center (NISC),

BC030301 Gaithersburg, Maryland;

BC030301 Web site: http://www.nisc.nih.gov/

Contact: nisc_mgc@nhgri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.C., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Graniter, S., Guan, X., Gupta, J., Haghghi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,
Maduro, Q.L., Masliello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgoun, C., Voigt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAP Plate: 65 Row: 9 Column: 11

This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis

This clone has the following problem: incomplete processing.

FEATURES

source

Location/Qualifiers
1. 1856
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5367539"
/tissue_type="Eye, retina, mouse strain C57Bl/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 386 a 545 c 472 g 453 t
ORIGIN

Query Match 19.9%; Score 687; DB 11; Length 1856;

Best Local Similarity 70.6%; Pred. No. 7.7e-148;

Matches 1228; Conservative 0; Mismatches 400; Indels 111; Gaps 19;

```

QY 1721 GGTATGGCTTTGAGGGTGTGATCTCGACGATCTATGGCATGGAGGAGACCTGACAT 1780
DB 154 GGGATGTGCTTTGAGGGGTGTGATCTCGACCATCTATGATGGAGCGAGGACCTGACCT 213
QY 1781 GTTTAGAGAACGCTGCGCGTTCGGGAGGACAGAGCATCTCCGAGCGCTGGATGG 1840
DB 214 GCTTAGTAGACAGTGCCTTCGGGACCCAGACCAATATTTGGGTGAGTGTAGTAG 273
QY 1841 AGGATGCCAAGCTCTACATGACATCTCTGTGCTTTGGGATCTTTCTCTAGCCCTGGGC 1900
DB 274 AGGAGGCCAAACTACATGACATCTCTGTGCTTGGGCAATTTCTCTTCCCTGGCGG 333
QY 1901 TGCTGGCTACTGTGCTGTGCTTACCGGTCAGGTCAGAGATAGAGGCTTCCGCCAG 1960
DB 334 TGCTGGCATACCTTGTGCTCGGTACCGGTCAGGTCAGGTCAGAGATAGAGGCTTCCCTA 393
QY 1961 CCTGTACCCAGCCCTGACAGAGAACGCCAGCCAGCTCCAGCCCTTTGGGACTGTTTAA 2020
DB 394 CCTCTC-----TGGSCAGGAAGGCCAGTCCAGCCCTTTGGGCTGTTTAA 443
QY 2021 CCTATAGACTGGGACTGTTCTCGGGGGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCT 2080
DB 444 ACTTGT--ACTTTGGGCAATGTTGCTGTGTGAGCTGCTCTCTCTCTCTCTCTCTCTCTCT 501
QY 2081 CAGCTGCTGCTC---GACATGCTCTCCAGCTGGGCTCTGGAGTGGGGGCTCCAGCC 2137
DB 502 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 561
QY 2138 CTCCCACTATGCTCCAGAGTCTTCCCAAGTTGATGGGTTGTAGTCTCTCTCTCTCTCTCT 2197
DB 562 CTCCCGTC-RAGGCCAGGAGTCTTCCCAAGTTGATGAGTGTAGTCTCTCTCTCTCTCTCT 620
QY 2198 TCTCAACACCTGATGCAAGACTACTGGGAGGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCT 2257
DB 621 TCTCAACACCTGATGCAAGACTGCTGGAGGTTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT 679
QY 2258 CCTCTCTGCTGCTGCTGCTGGAGGCTCTAGGCTCTAGGCTCTAGGCTCTAGGCTCTAGGCT 2317
DB 680 CCTCTCTGCTGCTGCTGCTGGAGGCTCTAGGCTCTAGGCTCTAGGCTCTAGGCTCTAGGCT 739

```

```

QY 2318 AAAGTGGCCCTCTGCGGGTCCCAACACACAAAGTGTGTTGTAACCTGGCTGCTATAAG 2377
DB 740 AGATTGACTCCCTCTGGGTCCCGCCACAC-----TCTGCTATATCA 782
QY 2378 GTTGAGTTCAGGGCTGGGGCTGGTGGAGTCCCACTGGAAGTCCCATTTAGTATGGA 2437
DB 783 GTTGAGTTCAGGGCTGGGGCTGGTGGGTCTCCAGGAGTCCCATCATGACATTTGA 842
QY 2438 AATGACAGGGAAGACTCTGGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2497
DB 843 ATGGAGCAGGAGGAGTCTGGAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
QY 2498 ACCCTGGCTGACTTGGCAATCTCCAGCAGACAGAGTGGGTTTCTCTCTAGTACCA 2557
DB 901 ACTTGGCTAATTTGGACTATCTGTTGGACAGAGCGGGCTTCTCTCTGCGGCCACCA 960
QY 2558 CTCCCAATCTCTGGGATTTGGAGCGCTGGGGTGTGGGATGCCCCCATCTCCCTCTCCCAT 2617
DB 961 CTCCCAATCTCTGGAGATTGGAGAGCATAGGATGT-----CCCGGCCCTTCCCTA 1012
QY 2618 CACCTTTGGTGGGCGAGGCGCTGGTGCACCTGTGCAATATATGCTGTCTCTCTCTCTCT 2677
DB 1013 AGCTTTGGTCTGGATAGACCTCAACATCACTCTGCAATATATGCTGTGCTCTCTCTCT 1072
QY 2678 ACCTGCCACTGGAATGAGAAATGCACATTTATCTGGCGGGGTGAGTGGGGGAAGAC 2737
DB 1073 ATCGGCCACAGGACTGGAGAAATGCACATTTATCT-----GGGGGTGAGGCGGGAGCC 1128
QY 2738 CCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2797
DB 1129 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1188
QY 2798 CCGAGTGTACAGGACATACATAGAGACAGCCATCTCAGCCCTACACACTTGGCATCTCC 2857
DB 1189 ATCACATACAGGACACAAAGCGCGGCAATCTCAGCCCTACACAGCAGCATCTCC 1248
QY 2858 CTACAGCACAGAGGAAGTGTGTTGGCATCTCTGGTGGCGGTGCTGTTGGGAGGA 2917
DB 1249 TTGTGGCTCAGAGAGAGTGGCGTGCA-----GGGAG 1283
QY 2918 CAGTGGCAACCTCTCTCTGGGATCCATGTTGGAGACTCTAAGGATTAAGCTGGTGTG 2977
DB 1284 CAGTGGCAACCTCTCTCTGGCAATCTCCACATTTG--GAGGCTAAGAAATGAGACTGGTGTGA 1342
QY 2978 CCCAGGTGTCTACAGAACTGCAGGTCTCTACCCCAAGTCTTCCCTCTCTCTCTCTCTCT 3037
DB 1343 CCCAGTGTCTACAGGTCTCGGAGTGTCTATCTCCAGTCTTCTCTCTCTCTCTCTCTCTCT 1402
QY 3038 GGGTGGCACAGGCACTAGATCTCTGGAGTTCCAGAACCAACAGCAACAGCAACAGGG 3097
DB 1403 CGGG-----GTGGCGTGGCGTCTGGAGTCAAGACCCA--ACAAGCACAAGATGGG 1454
QY 3098 CATAGTTGGCTTTGGGCACTGCCACGAGCGCTCTCTTTTGTGCTCCATGCTGGGATC 3157
DB 1455 CATAGTGGCTGGCGGCACTGCTCTCCCATAGGCTCTCTTTGTGCGCCAAAGCTGACAC- 1513
QY 3158 TTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3216
DB 1514 -TCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1572
QY 3217 TCCACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3276
DB 1573 TTCA-TCTTCGGTTCAGAGTGTCTCTGGACATCAGAGCGGGGCTGGACATCAG 1631
QY 3277 GTGCGACACACTCAGAGCATGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3336
DB 1632 GTGCGACACAC-----CTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1683
QY 3337 TGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3396
DB 1684 TGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1728
QY 3397 TCTCTCAACCTGCCACAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3455

```

```

Db 1729 TCCTCTGTATCTCCCGCAGTCTCTCTCATTTACACAGAGA--CCAGTGTATTGTGGG 1785
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 2
BQ179452
LOCUS
DEFINITION
  UI-M-EW0-bw-e-17-0-UI.r1 NIH_BMAP_EW0 Mus musculus cDNA clone
  IMAGE:5703928 5', mRNA sequence.
ACCESSION
  BQ179452
VERSION
  BQ179452.1 GI:20354957
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 851)
AUTHORS
  NIH-MGC http://mgc.nci.nih.gov/.
TITLE
  National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. James Lin, University of Iowa
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LNL at:
  http://image.llnl.gov
  This clone was contributed by the Brain Molecular Anatomy Project
  (BMAP)
FEATURES
  Seq primer: pYX-5.
  Location/Qualifiers
    1..851
    /organism="Mus musculus"
    /strain="C57BL/6"
    /db_xref="taxon:10090"
    /clone_lib="IMAGE:5703928"
    /clone_lib="NIH_BMAP_EW0"
    /tissue_type="whole brain"
    /dev_stage="embryo 15.5 dpc"
    /lab_host="DH10B (TI phage resistant)"
    /note="Organ: Brain; Vector: pYX-Asc; Site_1: EcoR I;
    Site_2: Not I; The library was constructed according to
    Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
    1996. Denatured mRNA was size fractionated on a 1% agarose
    gel. First strand cDNA synthesis was primed with an
    oligo-dT primer containing a Not I site. Double stranded
    cDNA was size selected according to mRNA size fraction,
    ligated with EcoR I adaptor, digested with Not I, and then
    cloned directionally into pYX-Asc vector. The library tag
    sequence located between the Not I site and the polyA tail
    is CTGGCTGGAA. This library was created for the
    University of Iowa Mouse Brain Molecular Anatomy Project
    (BMAP). 'Gene Discovery in the Developing Mouse Nervous
    System', supported by National Institutes of Mental Health
    (NIMH), Hemin Chin, Ph.D., program coordinator."
  BASE COUNT 168 a 264 c 205 g 210 t 4 others
  ORIGIN
    Query Match 18.4%; Score 634.8; DB 14; Length 851;
    Best Local Similarity 88.9%; Pred. No. 5.9e-136;
    Matches 684; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
    QY 570 GGGCGCTGATCTGCTCCACAGGAGGAGCGCTCTCTCTGGCGGGGAGGAGCG 629
    Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    1 GGGACTGATCTGCTCTCACAGGAGGAGCTCTCTCTCTGGGAGGAGGAGCG 60
    QY 630 TCTGGCCATCGCTGGAGCTGTTCACAAACCGCGCTGTCTCTTTTCATGAGCCAC 689
    Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
    61 CTGGGCCATGCTCTGGAAGTGGTCAACACCGCGCTGTCTCTTTTCATGAGCCAC 120
    QY 690 CAGTGTCTGGATAGCGCCTCTTGTTCACAGTGGTGTCCCTCATGAAGTCCCTGGCACA 749

```

```

Db 121 CAGCGGTCTAGACAGCGCTTCTTGTTCCTCAAGTGGTGTCCCTCATGAAGTCCCTGGCCCA 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 750 GGGGGGCGGTACCACATCATCTGCACCATCCACCAGCCAGTGCACAGCTCTTTGAGATGTT 809
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TGGGGGCGCCACCGTCTCATCTGCACCATCCACCAGCCAGTGCACAGCTCTTTGAGATGTT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 810 TGACAAAGCTCTACATCCTGAGCCAGGCTCAGTGCATCTTCAAAGGAGTGGTCAACCACT 869
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TGACAAAGCTCTACATCCTGAGCCAGGCGCAATGCATCTTCAAAGGCGTGGTTACCACT 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 870 GATCCCTTCTAAAGGAGCTCGGCTTGCATTCGCCACCTACCAACCCGCTGACTT 929
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GATTCCCTTCTGAAGGGGCTTGGCTTGCATCTGCCACCTACCAACCCGCTGACTT 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 930 CATCATCGAGTGGCGCTCTGGCGAGTATGAGACCTGAACCCCATCTTGTTCAGGCGTGT 989
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CATCATTTGAGTGGCGCTCTGGAGAGTATGGAGACCTGAACCCCATCTTGTTCAGGCGTGT 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 990 GCAGAAATGGGCTGTGGCGCTATGGCTGAGAAAGAGCAGCCCTGAGAAGAACAGGTCCC 1049
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GCAGAAATGGGCTTTCACCATGGCCGAGAAAGAGCAGCCCTGGGAAGAAATGAGTCCC 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1050 TGGCCCATGCGCTCTTGTCTCCGGAAGTGGATGCCATTTGAAAGCCACACCTTTGGCCAC 1109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TGGCCCATGCGCTCTTGTCTCCGGAAGTGGATGCCATTTGAAAGCCACACATTTGGCCAC 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1110 CAGCACCTTCACACAGTCTGCACTCTCTTCAAGAGGACCTCTCTGCTCCATCTCAGGGA 1169
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 CAGCACCTTCACACAGTCTGCACTCTCTTCAAGAGGACCTCTCTGCTCCATCTCAGGGA 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1170 CAGCGTCTGACCCACCTACGCTTTCATGTCCTCCAGCTGGTATTGCGGTGCTCATCGGCT 1229
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 CAGCGTCTGACCCACTTTCGCTTTCATGTCCTCCAGCTGGTATTGCGGTGCTCATCGGCT 660
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1230 CTCTACCTGCATATTGGCGAGCATGCCACAGGCTTTCACAAACACCGGCTGCTCTT 1289
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CTCTACCTGCATATTGGCGAGTATGCCAGCAGAGTCTTCAACANCAACCGGCTTCTCTT 720
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1290 CTCTCTCATCTGCTCTCATGTCGCGCGCTCATGCGCACTGTGCTC 1338
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 CTCTCTCATCTGCTCTCTCTCATGTCGCGCTCATGCGCACTGTGCTC 769
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 3
BI731100
LOCUS
DEFINITION
  603351741F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5359019 5',
  mRNA sequence.
ACCESSION
  BI731100
VERSION
  BI731100.1 GI:15708113
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 778)
AUTHORS
  NIH-MGC http://mgc.nci.nih.gov/.
TITLE
  National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
  Unpublished (1999)
COMMENT
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: The Cepko Laboratory
  cDNA Library Preparation: Life Technologies, Inc.
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone Distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LNL at:
  http://image.llnl.gov
  Plate: LLAM11913 row: b column: 12
  High quality sequence stop: 762.
  Location/Qualifiers
    1..778
  FEATURES
    source

```

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5359019"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="Dh10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-SPORT6; site_1: NotI;
site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      157 a   249 c   179 g   193 t
ORIGIN

Query Match      17.5%; Score 603; DB 13; Length 778;
Best Local Similarity 89.1%; Pred. No. 1.3e-128;
Matches 67; Conservative 0; Mismatches 80; Indels 2; Gaps 2;

QY 690 CAGTGTCTGGATAGGCGCTCTTTTCCAAAGTGGTCTCCCTCATGAAGTCCCTGGCACA 749
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 14 CAGCGGTCTAGACAGCGCTCTTTTCCAAAGTGGTCTCCCTCATGAAGTCCCTGGCACA 73

750 GGGGGCGGTACCATCATCTGCACATCCACAGCCAGTGCACAGCTCTTTGAGATGTT 809
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
74 TGGGGCGCGCACCGTCTATCTGCACATCCACAGCCAGTGCACAGCTCTTTGAGATGTT 133

810 TGACAGCTCTACATCTCTGAG-CCAGGTCAGTGCATCTTCAAGGAGTGGTCCACCAAC 868
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 TGACAGCTCTACATCTCTGAGACAGGGCAATGCATCTTCAAGGGGTGGTTACCAAC 193

869 TGATCCCTCTATCAAGGAGTCTGGCTTGCTTCCCTCCCTACCTACCAACCCCGCTGACT 928
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 TGATCCCTCTATCAAGGAGTCTGGCTTGCTTCCCTCCCTACCTACCAACCCCGCTGACT 253

929 TCATCATGAGTGGCTCTGGCGAGTATGAGACCTGAACCCCATGTGTTCCAGGCTG 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 TCATCATGAGTGGCTCTGGAGAGTATGGAGACCTGAACCCCATGTGTTCCAGGCTG 313

989 TGCAGATGGCTGTGCGCTTATGGCTGAGAAGAGCAGCCCTTGAGAAGCAGGCTC 1048
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 TGCAGATGGCTTTCACCATGCGCGAGAGAGAGCAGCCCTGGGAAGATGAGCTCC 373

1049 CTGCCCCATGCCCTCTGTGCTCCGGAAGTGGATCCCATTTGAAGCCACACCTTTGCCA 1108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
374 CTGCCCCATGCCCACTTGCCCTCCGAGCTGATCTATTGAAGCCACACATTTGCCA 433

1109 CCAGCACCTTCACAGTCTGTGATCTCTTCAAGAGGACCTTCCTGTCCATCCTCAGG 1168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
434 CCAGCACCTTAACCCAGTCTGTGATCTCTTCAAGGAGGACTTTCTTGTCCATCCTCAGG 493

1169 ACAGGTCTTGACCCACCTACGCTTCATGTCCACCGTGTATTGGCGTCTCATCGGC 1228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
494 ACAGGTCTTGACCCACCTTGGCTTCATGTCCATCATGTGTCGATCGCGGTCTCATGGCC 553

1229 TCCTCTACCTGCATATGGGAGCATGCCAGCAAGTCTTCAACAAACCGGCTGCCTCT 1288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
554 TCCTCTACCTGCATATGGGAGCATGCCAGCAAGTCTTCAACAAACCGGCTGCCTCT 613

1289 TCCTCTCCATGCTTCTTCATGTTCCGCGCTTCATGCCAATGTGCTCACCTTCCCT 1348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
614 TCCTCTCCATGCTTCTTCATGTTCCGCGCTTCATGCCAATGTGCTCACCTTCCCT 673

1349 TAGAGATGGCGGTCTTCATGAGGAGCAGCTCAACTACTGGTACAGCCCTCAAGGCTATT 1408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
674 TAGAGATGGCGGTCTTCATGAGGAGCAGCTTGATCTACTGGTATATCTTAAAGCCTATT 733

1409 ACCTGCCAAGACCATGGCTGAGTGCCTTCAG 1443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
734 A-CTTCCGAGACCATGGCTGATGTGCCCTTCAG 767

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hd07h11"
/clone_lib="Human Retina cDNA (Un-normalized, unamplified)"
/tissue_type="Retina"
/lab_host="EMDH10B"
/notes="Organ: Eye; Vector: pSPORT1; Neural retina tissue
eye dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library in the
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life technologies) was constructed at
Pserve Biotechnology (Laurel MD) essentially following
the protocols of the SuperScript Plasmid System full
details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor [5'-pGACTAGTCTAGATCCGAGCGGCCGCC(T)15-3'
]. EST analysis was performed on the unamplified library
at the NIH Intramural Sequencing Center (NISC)."
```



```

|||||
241 ACTGAGGTGTACCCCAAGTCTCTCCCTCCCAAGCCAGGGTGGCACAGGCAC 300
|||||
3056 AGATCCCTGAGTTCAGGAACCAACACAGCAACACAGCCAGGGCATAAGTTGGCC 3115
|||||
301 AGATCCCTGAGTTCAGGACCAACACAGCAACACAGCCAGGGCATAAGTTGGCC 360
|||||
3116 ACTGCCACCCAGCGCCCTCTTTTGTGCTCCATGCTGGCATCTTCACTCCCTACCCCT 3175
|||||
361 ACTGCCACCCAGCGCCCTCTTTTGTGCTCCATGCTGGCATCTTCACTCCCTACCCCT 420
|||||
3176 CCCAGCCACTGCTGCTATTCATCAAACTTCTGTCCATGCTCCCTCCACTGTTCTATCAGCA 3235
|||||
421 CCCAGCCACTGCTGCTATTCATCAAACTTCTGTCCATGCTCCCTCCACTGTTCTATCAGCA 480
|||||
3236 GGTGGCCCTGGGCATCAGACAGCCTGCCCTGGGCACAGGTGGCAGACACTCAGAG 3295
|||||
481 GGTGGCCCTGGGCATCAGACAGCCTGCCCTGGGCACAGGTGGCAGACACTCAGAG 540
|||||
3296 CATGCTG 3303
|||||
541 CATGCTG 548

```

```

RESULT 5
BM925745
LOCUS
DEFINITION AGENCOURT_6710400 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5763981
5', mRNA sequence.
ACCESSION BM925745
VERSION BM925745.1 GI:19376124
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12816 row: k column: 22
High quality sequence start: 34
High quality sequence stop: 614.

```

FEATURES

Source

```

1..1045
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5763981"
/clone_lib="NIH_MGC_114"
/lab_host="PH10B"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
male brains, age range 23-27 yo. Library is oligo-dr
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.5 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invited). Research Genetics tracking code 019. Note:
this is a NIH_MGC Library."
BASE COUNT 185 a 334 c 367 g 159 t
ORIGIN

```

Query Match

```

Best Local Similarity 15.7%; Score 541.6; DB 14; Length 1045;
Pred. No. 2.5e-114;

```

```

Matches 617; Conservative 0; Mismatches 49; Indels 6; Gaps 4;
QY 7 ATGGCGGAGAACGCGCTGGAGCCGCTGGGCTGTGGACTAGGCGCGGGGCTGTGGCCATG 66
|||||
347 ATGGCGGAGAACGCGCTGGAGCCGCTGGGCTGTGGACTAGGCGCGGGGCTGTGGCCATG 406
|||||
67 GCCGTGACGCTGGAGGACGGGCGGAACCCCTGTGCTGACACACACCTGAAGAAGTGG 126
|||||
407 GCCGTGACGCTGGAGGACGGGCGGAACCCCTGTGCTGACACACACCTGAAGAAGTGG 466
|||||
127 GAGAACACATCACTGAAGCCCGAGCGCTTCTCCACCTGCCCAAGCGCTCAGCCGTGGAC 186
|||||
467 GAGAACACATCACTGAAGCCCGAGCGCTTCTCCACCTGCCCAAGCGCTCAGCCGTGGAC 526
|||||
187 ATCGAGTTCGTGAGCTGTCTTATTCGTCGGGAGGGGCGCTGTGGCGCAAAAGGGGT 246
|||||
527 ATCGAGTTCGTGAGCTGTCTTATTCGTCGGGAGGGGCGCTGTGGCGCAAAAGGGGT 586
|||||
247 TATAAGACCCCTTCTCAAGTGCCTCTCAGTAAATTTCTGCCGCGGAGAGCTGATTTGCCATC 306
|||||
587 TATAAGACCCCTTCTCAAGTGCCTCTCAGTAAATTTCTGCCGCGGAGAGCTGATTTGCCATC 646
|||||
307 ATGGGCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 366
|||||
647 ATGGGCCCTCAGGGCTGGCAAGTCTACATTCATGAACATCTTGGCAGGATACAGGAG 706
|||||
367 TCTGGAATGAAGGCGAGATCCTGTTAATGAAGGCCACGGAGCTGAGGACCTTCCCG 426
|||||
707 TCTGGAATGAAGGCGAGATCCTGTTAATGAAGGCCACGGAGCTGAGGACCTTCCCG 766
|||||
427 AAGATGTCCTGCTACATCATGCAAGTACATGCTGCTGCCACACCTCACGGT-GTTGGA 485
|||||
767 AAGATGTCCTGCTACATCATGCAAGTACATGCTGCTGCCACACCTCACGGTGTGGGA 826
|||||
486 ACCCATGATGCTCTGCTTAACCTGAATCTTACTGAGATCCCGATGTG-AATAACCATC 544
|||||
827 ACCCTGGATGGTCTGCTTAACCTGAAGCTTGAAGACCAAGAGGGGAAAAAGAGC 886
|||||
545 TCGTGACAGAGATCTGACGGCACCTGAGCTGCTGCTGCCACACGAGGACAGCC 604
|||||
887 TGGTGACGAGAACCTGACGGCACGGGGCCCGATGTCGGCTTCCCCCAGGACAGCC 946
|||||
605 TGCTCTCTGGCGG---CAGAGGAAGCGTCTGCCCATCCCTGGAGCTG-GTCAACAAC 660
|||||
947 TGGTCTTGGCGGGCAAGGAAGCTCTGTCGCCATCCCTGGAAGTGTGGCAAAAC 1006
|||||
661 CGCCCTGTGATG 672
|||||
1007 CGCGCGGGCAAG 1018
|||||
RESULT 6
AK018283
LOCUS
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430517004:hypothetical protein, full
insert sequence.
ACCESSION AK018283
VERSION AK018283.1 GI:12857918
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male olfactory brain cDNA to
mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
clone:6430517004.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2

```


cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10368 row: j column: 12
 High quality sequence stop: 627.
 High quality sequence stop: 627.

FEATURES

Location/Qualifiers
 1..974
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:4501355"
 /clone_lib="NIH_MGC_94"
 /tissue_type="retina"
 /lab_host="DH10B (phage-resistant)"
 /note="organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 3.3 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."
 199 a 312 c. 253 g 210 t

BASE COUNT
 ORIGIN
 Query Match 11.8%; Score 406.6; DB 12; Length 974;
 Best Local Similarity 84.7%; Pred. No. 4.1e-83;
 Matches 526; Conservative 0; Mismatches 89; Indels 6; Gaps 6;

QY 690 CAGTGGTCTGGATAGCGCTCTGTTCACAGTGTGTCCTCATGAAGTCCCTGGCACA 749
 DB 140 CAGCGGTCTAGACAGCGCTCTGTTCACAGTGTGTCCTCATGAAGTCCCTGGCACA 199
 QY 750 GGGGGGCCCTACCATCATCTGCACCATCCACCAGCCAGTGCACAGCTCTTTGAGATGTT 809
 DB 200 TGGGGGCCCGACCGTCATCTGCACCATCCACCAGCCAGTGCACAGCTCTTTGAGATGTT 259
 QY 810 TGCAAGCTCTACATCTGAGCCAGGGTCAATCTTCAAGAGTGTGCACCAACT 869
 DB 260 TGCAAGCTCTACATCTGAGCCAGGGTCAATCTTCAAGAGTGTGCACCAACT 318
 QY 870 GATCCCTCATTAAGAGGACTCGCTGCTGATTCGCCCTACCAACCCGCGTCACTT 929
 DB 319 GATTCCTCATTAAGAGGACTCGCTGCTGATTCGCCCTACCAACCCGCGTCACTT 378
 QY 930 CATCATCAGGTGGCTCTGGCGAGTATGGAGACCTGAACCCCATGTTCTCAGGCGCT 989
 DB 379 CATCATCAGGTGGCTCTGGCGAGTATGGAGACCTGAACCCCATGTTCTCAGGCGCT 438
 QY 990 CGAAGTGGCTGTGCGCTATGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1049
 DB 439 CGAAGTGGCTGTGCGCTATGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 497
 QY 1050 TCGCCCATGCGCTCTGCTGCTGCGAGTGTATCC-CAATTAAGAGAGAGAGAGAGAG 1108
 DB 498 TCGCCCATGCGCTCTGCTGCTGCGAGTGTATCC-CAATTAAGAGAGAGAGAGAGAG 557
 QY 1109 CCAGCAGCTCTACAGAGTCTGCTGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 1168
 DB 558 CCAGCA-CTTAAGAGAGTCTGCTGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 616
 QY 1169 ACAGGCTCTGACCACTACAGGTTATGTCACAGTGTGTTATGGCGTGTCTATCGGCC 1228
 DB 617 ACAGGCTCTGACCACTACAGGTTATGTCACAGTGTGTTATGGCGTGTCTATCGGCC 674
 QY 1229 TCCTCTACTGTCATATGGCGAGGATGCCAGAGGCTCTTCAACACAGCGGCTGCCTCT 1288
 DB 675 CCCTCTACTGTCATATGGCGAGGATGCCAGAGGCTCTTCAACACAGCGGCTGCCTCT 734
 QY 1289 TCCTCTACTGTCATATGGCGAGGATGCCAGAGGCTCTTCAACACAGCGGCTGCCTCT 1309
 DB 735 TCCTCTACTGTCATATGGCGAGGATGCCAGAGGCTCTTCAACACAGCGGCTGCCTCT

RESULT 10

BE742274
 LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BE742274
 601575692F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3836615 5',
 mRNA sequence.
 BE742274
 BE742274.1 GI:10156266
 EST.
 human.
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 964)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM520 row: h column: 24
 High quality sequence stop: 681.

Location/Qualifiers
 1..964

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3836615"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(C). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 181 a 303 c 277 g 203 t

ORIGIN

Query Match 11.5%; Score 397.4; DB 12; Length 964;
 Best Local Similarity 78.0%; Pred. No. 5.5e-81;
 Matches 492; Conservative 0; Mismatches 136; Indels 3; Gaps 1;

QY 1234 TACCTGTCATATTGGCGACGATGCCAGAGTCTTCAACAACACCGCTCTCTCTTC 1293
 DB 1 TACTTGGGATCGGGNACGAGCCAGAGAGGCTTTCAGCAACTCCGGCTCTCTCTTC 60

QY 1294 TCATGCTCTTCTCATGTTTCGCGCCCTCATGCGCAACTGCTCACTACCTTCCCGCTTAGAG 1353
 DB 61 TCCATGCTCTTCTCATGTTTCGCGCCCTCATGCGCAACTGCTCACTACCTTCCCGCTTAGAG 120

QY 1354 ATGGCGGCTCTCATGAGGAGCAGCTCACTACTGTCAGCCTCAAGCGTATTACCTG 1413
 DB 121 ATGGGAGTCTTCTTCGGGAACACCTGAACACTACTGTCAGCCTCAAGCGTATTACCTG 180

QY 1414 GCCAAGACCATGGCTGACGTGCCCTTTTCAGGTGGTGTGTCGGGTGCTTACTGACGACT 1473
 DB 181 GCCAAGACCATGGCAGACGCTGCCCTTTTCAGATCATGTTCCCGAGTGCCTTACTGACGATC 240

QY 1474 GTGTACTGATGACGGGCGACCGCTGAGACAGCGCTTCTGCTCTTCTTTCAGCGCTG 1533
 DB 241 GTGTACTGATGACGTCGACGCGCTCCGACGCGCTTGTGTGTGTTGTCGCGCTG 300

QY 1534 GCCACCGCCACCGCTTGGTGGCCCACTTTTGGGCTCTGCTGATCGGAGCTGCTTCCAC 1593
 DB 301 GGCACCATGACCTCCCTGGTGACAGCTCCCTGGGCTCTGATCGGAGCCCGCTCCACG 360

Qy	1594	TCCCTACAGTGGCCACTTTTGTGGGCCAGTTACCGCCATCCCTGHCCTCTGTGTTCTCC	1653
Db	361	TCCCTGCAGTGGCCACTTTTGTGGGCCAGTGCACAGCCATCCGGTGTCTCTGTTCTCG	420
Qy	1654	GGCTTCTTTGTCACTTCAAGACCATCCGCATTAAGTCAATGGAGCTCCTATCTCTCC	1713
Db	421	GGTTTCTTGTCACTTCGACACCATCCACACGTACCTACATGGATGTCCTACATCTCC	480
Qy	1714	TATGTCAGGATGGCTTTGAGGTGTGATCCTGACGATCTATGCGATGGAGCGAGGAGAC	1773
Db	481	TATGTCAGGATGGGTTCCGAAGGGGTATCCTCTCCATCTATGGCTTAGACACGGGAAGAT	540
Qy	1774	CTCACATGT--TTAGAGGAACCTGCCGTTCCGGGAGCCACAGAGCATCTCTCCGAGCG	1830
Db	541	CTGCATGTGACATCAGCAGAGCTGCCACTCCAGAAGTCGGAGGCCATCTCTCGGGAG	600
Qy	1831	CTGGATGTGGAGGATGCCAAGCTCTACATGG	1861
Db	601	CTGGAGCTGGAAATGCCAAGCTGTACCTGG	631

30068
 JUL 11

LOCUS	BQ930068	885 bp	linear	EST 20-AUG-2002
DEFINITION	AGENCOURT_8931968	NIH_MGC_94	Mus musculus	cDNA clone IMAGE:6465973
				5', mRNA sequence.

ACCESSION	BQ930068
VERSION	BQ930068.1
GI	22345099

KEYWORDS
SOURCE
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 885)
 NIH-MGC <http://mgc.ncbi.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 COMMENT

Email: cgabbs@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:

FEATURES SOURCE

BASE COUNT	155 a	284 c	225 g	221 t	ORIGIN
Query Match					
Best Local Similarity					11.1%; Score 383.2; DB 14; Length 885;
Matches 648; Conservative					72.6%; Pred. No. 1e-77;
					Mismatches 0; Indels 46; Gaps 10;
QY	1981	GCAGGAAGCCCCCAGTCCAGCCCTTTGGACCTGTTTTAACCTTATAGACTTGGGGCAGTC	2040		
Db	27	GCAGGAAGCCCCAGTCCAGCCCTTTGGGGCTGTTTTAAACTTGT--ACTTGGGCATTG	84		
QY	2041	GTTCCTGTGGGGGGCTATCTCTCCTCCCTTGGCTCTCCACAGGCTGGCTGTCT--GGAC	2097		

[illegible]

RESULT 12	
BF159563	
LOCUS	879 bp mRNA linear EST 30-OCT-2000
DEFINITION	G01769059F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3988239 5'
ACCESSION	BF159563
VERSION	BF159563.1
KEYWORDS	GI:11039466
SOURCE	EST.
ORGANISM	house mouse
	Mus musculus
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
REFERENCE	1 (bases 1 to 879)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Db	604	CACGGCTCCTGGCTCCAAAGTGGTGCTCTGATGAGAGGACTGGGCCACGGGGTGGGGCT	663
Qy	761	CCATCATCTGCACCATCCACCCAGCTCTTTGAGACTTTTGAACAAGCTCT	820
Db	664	CCATCGTCTGTACCATCCAAACGACGTGCAAGCTCTTAGAGCTCTTTGACCAGCT-T	722
Qy	821	ACATCTGTAGCCAGGCTCAGTGCATCTTCAAAGGAGTGTCACCACTGATCCCCATTC	880
Db	723	ATCTCCTAAGTCAGGAAATGCGTTTACCGGGGAAGCGTTTCCATCTCGTGCCGGATC	782
Qy	881	TAAAGGACTCGGCTTGCAATTGCC	904
Db	783	TGAGGGACCGGGTCTGAAATGGCC	806

RESULT 13					
B06468					
LOCUS	B06468	688 bp	DNA	linear	GSS 13-JUL-1996
DEFINITION	CSRL-80H4-u CSRL flow sorted Chromosome 11 specific cosmid Homo sapiens genomic clone CSRL-80H4, DNA sequence.				

SOURCE: Human: Homo sapiens
ORGANISM: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE: 1 (bases 1 to 688)

TITLE
Genomic Sequence Sampled Map of Chromosome 11
JOURNAL
Unpublished (1996)
COMMENT
Contact: Evans GA, Shane Probst
McDermott Center for Human Growth and Development
University of Texas Southwestern Medical Center At Dallas
5323 Harry Hines Blvd, Dallas TX 75235-8591

FORTRANCED
 FORWARD: CTCATTCTGTTCTGATTTC
 BACKWARD: CCACACACATTTACACATTG
 Seq primer: T7
 Class: cosmid ends
 High quality sequence stop: 688.

```

1: 800
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSRL-80h4"
/clone_lib="CSRL flow sorted Chromosome 11 specific
cosmid"
/sex="female"
/cell_type="chimeric hamster somatic cell hybrid"
/notes="Vector: scOs-1; Human Chromosome 11 specific cosmid
library prepared from flow sorted human Chromosome 11
derived from Chinese Hamster Ovary (CHO) monochromosomal
somatic cell hybrid, J1"
BASE COUNT 148 a 190 c 128 q 185 t 37 others

```

Query Match	10.7%	Score 369;	DB 17;
Best Local Similarity	94.3%	Pred. No. 1.7e-74;	Length 688;
Matches 378;	Conservative	0;	Mismatches 23;
			Indels 0;
			Gaps 0;

Db 6 TNNNTTCTCGGANTTCGGCAATTNCAAAAGCACACACGGGCATTAAGTTGNCCTTGGC 65

Qy 3115 CACTGCCACCCAGGGCCCTCCTTTTGTCCTCCATGCTGGGCATCTTCACTCCCCTACCCCT 3174

Db 66 CACTGCCACCCAGCGCCCTCTTTGTGCTCCATGCTGGCATCTTCACTCCCTACCCCT 125
 QY 3175 TCCCGAGCCAGCTGCTCATTCACAACTTGTGTCATGTCCTCCACTGTTCTTATCAGC 3234
 Db 126 TCCCGAGCCAGCTGCTCATTCACAACTTGTGTCATGTCCTCCACTGTTCTTATCAGC 185
 QY 3235 AGTGGCCCTGGCATCAGACAGCTGCTCCCTGGGCACCAAGTGGCAGACACACTCACA 3294
 Db 186 AGTGGCCCTGGCATCAGACAGCTGCTCCCTGGGCACCAAGTGGCAGACACACTCACA 245
 QY 3295 GCATGCTGGCTTCTGCTGGTGCAGGCTCATTCGCTGCTGATTTCCCTCCCGCAG 3354
 Db 246 ACATGCTGGCTTCTGCTGGTGCAGGCTCATTCGCTGCTGATTTCCCTCCCGCAG 305
 QY 3355 GGTCTATTTTCCCTTTTCTGTCACATCCCTGCTACCTGCTTACCTCCCTCCCGCACA 3414
 Db 306 GGTCTATTTTCCCTTTTCTGTCACATCCCTGCTTACCTCCCTCCCGCACA 365
 QY 3415 GATTTCTTCTATCACACAGGATGCCAGTTGATTTGTGGG 3455
 Db 366 GATTTCTTCTATCACACAGGATGCCAGTTGATTTGTGGG 406

RESULT 14

B0769103
 LOCUS
 DEFINITION
 UI-M-FCO-byo-f-19-0-UI.r1 NIH-BMAP_FCO Mus musculus cDNA clone
 IMAGE:5720466 5', mRNA sequence.

ACCESSION
 VERSION
 B0769103.1 GI:21977577
 EST.
 SOURCE

ORGANISM
 Mus musculus
 house mouse.
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 The following repetitive elements were found in this cDNA
 sequence: 2-55, >MTDBLTR/MaLR (matched complement)
 Seq primer: pYX-5.
 Location/Qualifiers

FEATURES
 source
 1. .625
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5720466"
 /clone_lib="NIH-BMAP_FCO"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /note="organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
 Site_2: Not I; The library was constructed according to
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured mRNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was size selected according to mRNA size fraction,
 ligated with EcoR I adaptor, digested with Not I, and then
 cloned directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail

, is TGAGAGAGCC. This library was created for the
 University of Iowa Mouse Brain Molecular Anatomy Project
 (BMAP). 'Gene Discovery in the Developing Mouse Nervous
 System', supported by National Institutes of Mental Health
 (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 127 a 207 c 142 g 145 t 4 others
 ORIGIN
 Query Match 10.6%; Score 367.4; DB 14; Length 625;
 Best Local Similarity 89.8%; Pred. No. 3.8e-74;
 Matches 403; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

QY 690 CAGTGGTCTGGATAGCGCTCTGTTTCCAACTGGTGTCCCTCATGAAGTCCCTGCACCA 749
 Db 178 CAGCGGTCTAGACAGCGCTTCTGTTTCCAACTGGTGTCCCTCATGAAGTCCCTGCACCA 237
 QY 750 GGGGGCGGTAGCATCATCTGCACCATCCACAGCCAGTGCACCAAGCTCTTTGAGATGTT 809
 Db 238 TGGGGCGCGCACCGTCACTGTCACCATCCACAGCCAGTGCACCAAGCTCTTTGAGATGTT 297
 QY 810 TCACAAGCTCTACATCCTTGAGCCAGGTCAGTGCATCTTCAAGAGAGTGGTACCAACCT 869
 Db 298 GGACAAGCTCTACATCCTTGAGCCAGGGAATGCACTCTTCAAGGCGGTGGTTACCAACCT 357
 QY 870 GATCCCTCTATTAAGGACTCGGCTTGCAATGCCCCACCTACCAACCCGGCTGACTT 929
 Db 358 GATTCCTCTATCTNAAGGCGCTTGGCTTGCACCTGCCACCTACCAACCCGGCTNACTT 417
 QY 930 CATCATGAGGTGGCTCTGGCGAGTATGGAGACCTGAACCCCATGTTGTTTCAGGCGTGT 989
 Db 418 CATCATTNAGGTGGCTCTGGAGAGATATGGAGACCTGAACCCCATGTTGTTTCAGGCGTGT 477
 QY 990 GCAGAAATGGGCTGTCGGCTATGGCTGGAAGAAGAGAGAGCCCTGAGAGAGAGAGTCC 1049
 Db 478 GCAGAAATGGGCTTTCGACCATGCGCGAGAAGAAGAGAGAGCCCTGGAAGAATGAGCTCCC 537
 QY 1050 TGCCCCATGCCCTCTGCTGCTCCGGAAGTGGATCCCATTTGAAAGCCACACCTTTGGCCAC 1109
 Db 538 TGCCCCATGCCCTCTGCTGCTCCGGAAGTGGATCCCATTTGAAAGCCACACCTTTGGCCAC 596
 QY 1110 CAGCACCTTCACACAGTCTGTCATCTCT 1138
 Db 597 CAGCACCTTCACACAGTCTGTCATCTCT 625

RESULT 15
 BE262142
 LOCUS
 DEFINITION
 BE262142
 mRNA sequence.
 601151887F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3508053 5',
 BE262142
 mRNA sequence.
 647 bp mRNA linear EST 26-OCT-2000
 BE262142.1 GI:9134906
 VERSION
 BE262142.1
 KEYWORDS
 EST.
 SOURCE
 human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 647)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM188 row: n column: 22
 High quality sequence stop: 644.
 Location/Qualifiers
 1. .647
 source


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3508053"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB/; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
140 a 171 c 204 q 132 t

```

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Query Match	10.4%;	Score 360.2;	DB 10;	Length 647;
Best Local Similarity	73.6%;	Pred. No. 1.8e-72;		
Matches 472;	Conservative 0;	Mismatches 168;	Indels 1;	Gaps 1;
Qy	121	AAGGTGAGAACACATCACTGAAGCCACCGCTTCTCCACCTGCGCCACAGCGCTCAGCC	180	
Db				
Qy	1	AAAGTAGATAATAACCTCAGCGAAGCCACCGCTTCTCCTCTTGCTCGAGGGCGAGCT	60	
Db				
Qy	181	GTGGACATCAGATTGCTGGAGCTGTCTTATTTTCGTGCGGAGGGGGCCCTCTCTGCGCAAA	240	
Db				
Qy	61	GTGNACATTAATTCAGGGACCTTCTTATTCGGTTCCTGAAGACCCCTGGTGAGGAGG	120	
Db				
Qy	241	AGGGTTTATAAGACCCCTTCTCAAGTGCTCTCAGTAGTAATTTCTGCCGCCGGAGCTGATT	300	
Db				
Qy	121	AAAGGATACAGACCCCTCCTGAAGGAATTTCCGGGAAGTTCAATAGTGTGAGTTGGTG	180	
Db				
Qy	301	GGCATCATGGCCCTCAGGGCTTGGCAATTCATATCATATGACATCTTTGGCAGGATAC	360	
Db				
Qy	181	GCCATTATGGTTCCTTCGGGGCGGGAAGTCCACGCTGATGAACATCTCTGGCTGGATAC	240	
Db				
Qy	361	AGGAGTCTGGAATGAAGGGGCAGATCCTGGTTAATGGAAGGCCACCGGAGCTCAGGACC	420	
Db				
Qy	241	AGGGACGGGCATGAAGGGGGCGTCTCATCAACGGCTGCCCGGGACCTCGCGTGC	300	
Db				
Qy	421	TTCCGCAAGATGTCTCTACATCATCAAGATGACATGCTGTCTGCCGACCTCACGGTG	480	
Db				
Qy	301	TTCCGGAAGTGTCTCTACATCATCGAGATGACATGCTGTCTGCCATCTCCTACTGTG	360	
Db				
Qy	481	TTGGAACCATGATGTGTCTGTCTGAACCTGAATCTTTACTGAGAAATCCCGATGTGAAAAAC	540	
Db				
Qy	361	CAGAGGCCATGATGTTGTCTGGCACATCTGAAGCTTCAGGAGAAGGATGAAGGCAGAAGG	420	
Db				
Qy	541	GATCTCTGACAGATCTTCAGGGCACTGGGCTGATGTCTGTCTCCACACAGGAGACA	600	
Db				
Qy	421	GAATGTCTAAGAGATACCTGACAGCGCTGGGCTTGCTGTCTTTCGCGCAACACCGCGACC	480	
Db				
Qy	601	GCCTGTCTCTCTGGCGGACAGGAAGCGTCTGGCCATCGCCCTGGAGCTGGTCAACAAC	660	
Db				
Qy	481	GGGAGCTGTCAAGTGTGTAGCGCAAGCGCTTGCCCATCGCGCTGGAGCTGGTGAACAAC	540	
Db				
Qy	661	CCGCTCTCATGTTCTTTGTATGAGCCACCAGTGGTCTGGATAGCGCTCTTGTGTTCCAA	720	
Db				
Qy	541	CCTCCAGTCATGTTCTTCGATGAGCCCAACAGCGG-CTGGACAGCGCTCTCTGCTTCCAG	599	
Db				
Qy	721	GTGGTGTCCCTCATGAAGTCCCTGGGCACAGGGGGCCGTAC	761	
Db				
Qy	600	GTGGTCTCGTGTGAAGGGCTCGCTCAAGGGGGTCTGCC	640	
Db				

Search completed: June 16, 2003, 17:39:27
Job time : 4545 secs

[illegible]

Db	1321	CTCATGCCAACTGTGTCTACCTTTCCCTCTTAGAGATGGCGTCTTTCATGAGGGAGCACCTC	1381
Qy	1381	AAC TACTGGTACAGCCTCAAAGCGTATTACTTGGCCAAAGACCATTGGCTGACGTGCCTTTT	1440
Db	1381	AACTACTGGTACAGCCTCAAAGCGTATTACTTGGCCAAAGACCATTGGCTGACGTGCCTTTT	1440
Qy	1441	CAGTGGTGTGTCCGGTGGTCTACTCGACGATTTGTGTACTGATGATGAGGCCACGCCGCT	1500
Db	1441	CAGTGGTGTGTCCGGTGGTCTACTCGACGATTTGTGTACTGATGATGAGGCCACGCCGCT	1500
Qy	1501	GAGACAGCCGCTTCTGCTCTTCTCAGCCCTGGCCACGCCACCGCCTTGGTGGCCAA	1560
Db	1501	GAGACAGCCGCTTCTGCTCTTCTCAGCCCTGGCCACGCCACCGCCTTGGTGGCCAA	1560
Qy	1561	TC TTTGGGCTGTGATCGGAGCTGCTTCCAACTCCCTACAGGTGGCCACTTTGTGGGC	1620
Db	1561	TC TTTGGGCTGTGATCGGAGCTGCTTCCAACTCCCTACAGGTGGCCACTTTGTGGGC	1620
Qy	1621	CCAGTTACC GCCATCCCTGTCTTGTCTCCGGCTTCTTGTTCAGCTTCAAGACCATC	1680
Db	1621	CCAGTTACC GCCATCCCTGTCTTGTCTCCGGCTTCTTGTTCAGCTTCAAGACCATC	1680
Qy	1681	CCCACTTACCTGCAATGGAGCTCCTATCTCTCTATGTTCAGGTATGCTTTGAGGCTGTG	1740
Db	1681	CCCACTTACCTGCAATGGAGCTCCTATCTCTCTATGTTCAGGTATGCTTTGAGGCTGTG	1740
Qy	1741	ATCTGACGATCTATGGCATGGAGGAGACTGACATGTTTAGAGGAACCTGCCCG	1800
Db	1741	ATCTGACGATCTATGGCATGGAGGAGACTGACATGTTTAGAGGAACCTGCCCG	1800
Qy	1801	TTCCGGGAGCCACAGAGCATCCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATG	1860
Db	1801	TTCCGGGAGCCACAGAGCATCCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATG	1860
Qy	1861	GACTTCTGTCTTTGGGCATCTTCTTCTCCTAGCCCTGGGGCTGCTGGCCTACCTTGTGCTG	1920
Db	1861	GACTTCTGTCTTTGGGCATCTTCTTCTCCTAGCCCTGGGGCTGCTGGCCTACCTTGTGCTG	1920
Qy	1921	CGTTACCGGCTCAAGTCAGAGATAGAGCTTGGGACTGTTTAACTTATAGACTTTGGGCAC	1980
Db	1921	CGTTACCGGCTCAAGTCAGAGATAGAGCTTGGGACTGTTTAACTTATAGACTTTGGGCAC	1980
Qy	1981	GCAGGAAGCCCCAGTCCCAAGCCCTTTGGGACTGTTTAACTTATAGACTTTGGGCAC	2040
Db	1981	GCAGGAAGCCCCAGTCCCAAGCCCTTTGGGACTGTTTAACTTATAGACTTTGGGCAC	2040
Qy	2041	GTTCCTGGGGGGCTATCTCTCTCCCTTGGCTCCTCCACAGGCTGGCTGTCGGA	2100
Db	2041	GTTCCTGGGGGGCTATCTCTCTCCCTTGGCTCCTCCACAGGCTGGCTGTCGGA	2100
Qy	2101	GCTCCAGCCTGGCTGTGGAGTGGGGCTCCAGCCCTCCACACTATGCCAGGAGTCT	2160
Db	2101	GCTCCAGCCTGGCTGTGGAGTGGGGCTCCAGCCCTCCACACTATGCCAGGAGTCT	2160
Qy	2161	TCCCAAGTTGATCGGTTTGTAGCTTCTCCTCCTACTCTCTCCAAACCTGCATGCA	2220
Db	2161	TCCCAAGTTGATCGGTTTGTAGCTTCTCCTCCTACTCTCTCCAAACCTGCATGCA	2220
Qy	2221	CTACTGGAGGCTGTGCTCTCTCTCTGCCCATTGGCACCCCTCTCTGCTGTCTGCTGG	2280
Db	2221	CTACTGGAGGCTGTGCTCTCTCTCTCTGCCCATTGGCACCCCTCTCTGCTGTCTGCTGG	2280
Qy	2281	AGCCCTAGGCTCTTAGGGCCCACTTACAACCTGACCAAGTGGCCCTCTCTGGGGTCC	2340
Db	2281	AGCCCTAGGCTCTTAGGGCCCACTTACAACCTGACCAAGTGGCCCTCTCTGGGGTCC	2340
Qy	2341	CCACCACAAAGTGTGTTAACTGGGCTGCTATAAGTTTGAAGTTCCAGGGCTGGGCC	2400
Db	2341	CCACCACAAAGTGTGTTAACTGGGCTGCTATAAGTTTGAAGTTCCAGGGCTGGGCC	2400
Qy	2401	TGTTGGAGTCCACTGGAAGTCCCATTTATGGATGTTGAAATGGACAGGAGGACTCTGA	2460
Db	2401	TGTTGGAGTCCACTGGAAGTCCCATTTATGGATGTTGAAATGGACAGGAGGACTCTGA	2460

|||||
781 CAGCCAGTGCCAAAGCTCTTTGAGATGTTGACAGCTACATCCTGAGCCAGGTCAG 840 Db
|||||
841 TGCATCTTCAAGAGAGTGTGTCACCAACCTGATCCCGTATCTAAAGGACTCGGCTGCAT 900 QY
|||||
841 TGCATCTTCAAGAGAGTGTGTCACCAACCTGATCCCGTATCTAAAGGACTCGGCTGCAT 900 Db
|||||
901 TGCCCCACCTACCAACCCGGCTGACTTCATCATCGAGGTGGCCTCTGCGCAGATGGA 960 QY
|||||
901 TGCCCCACCTACCAACCCGGCTGACTTCATCATCGAGGTGGCCTCTGCGCAGATGGA 960 Db
|||||
961 GACCTGAACCCCATGTTGTTAGGGCTGTGCAAGTGGGCTGTGCGCTATGGCTGAGAAG 1020 QY
|||||
961 GACCTGAACCCCATGTTGTTAGGGCTGTGCAAGTGGGCTGTGCGCTATGGCTGAGAAG 1020 Db
|||||
1021 AAGAGCAGCCCTGAGAGAACAGGTCCCTGCGCCCATGCGCTTGTGCTCCGGAAGTG 1080 QY
|||||
1021 AAGAGCAGCCCTGAGAGAACAGGTCCCTGCGCCCATGCGCTTGTGCTCCGGAAGTG 1080 Db
|||||
1081 GATCCCATGAAAGCCACACCTTTGCCACGACCCCTCACACAGTTCTGCACTCTCTTC 1140 QY
|||||
1081 GATCCCATGAAAGCCACACCTTTGCCACGACCCCTCACACAGTTCTGCACTCTCTTC 1140 Db
|||||
1141 AAGAGACCTTCTGTCCATCCTCAGGACACGGTCTGACCCACCTACGGTTCTATGTCC 1200 QY
|||||
1141 AAGAGACCTTCTGTCCATCCTCAGGACACGGTCTGACCCACCTACGGTTCTATGTCC 1200 Db
|||||
1201 CAGGTGTTATTTGGGGTGTCTATCGGCTCTCTACCTGATATTTGGGACGATGCCAGC 1260 QY
|||||
1201 CAGGTGTTATTTGGGGTGTCTATCGGCTCTCTACCTGATATTTGGGACGATGCCAGC 1260 Db
|||||
1261 AAGGTCTTCAACAAACACCGGCTGCTCTTCTCTCCATGCTGCTCATGTTTCGCGCC 1320 QY
|||||
1261 AAGGTCTTCAACAAACACCGGCTGCTCTTCTCTCCATGCTGCTCATGTTTCGCGCC 1320 Db
|||||
1321 CTCATGCCAATGTGCTCACCCTTCCCTTTAGAGATGGCGGTCTTCATGAGGAGACCTC 1380 QY
|||||
1321 CTCATGCCAATGTGCTCACCCTTCCCTTTAGAGATGGCGGTCTTCATGAGGAGACCTC 1380 Db
|||||
1381 AACTACTGTACAGCTCAAAAGCTATTACCTGGCCAGACCATGCTGACGTGCCCTTT 1440 QY
|||||
1381 AACTACTGTACAGCTCAAAAGCTATTACCTGGCCAGACCATGCTGACGTGCCCTTT 1440 Db
|||||
1441 CAGGTGTTGTGCGGTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCT 1500 QY
|||||
1441 CAGGTGTTGTGCGGTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCT 1500 Db
|||||
1501 GAGACAGCCGCTTCTGCTCTCTCAGGCTGGCCACCGCCCTTGGTGCCCAA 1560 QY
|||||
1501 GAGACAGCCGCTTCTGCTCTCTCAGGCTGGCCACCGCCCTTGGTGCCCAA 1560 Db
|||||
1561 TCTTTGGGGTGTGATCGGAGTGTCTTCCAACTCCCTACAGGTGGCCACTTTTGTGGGC 1620 QY
|||||
1561 TCTTTGGGGTGTGATCGGAGTGTCTTCCAACTCCCTACAGGTGGCCACTTTTGTGGGC 1620 Db
|||||
1621 CCAGTTACCGCCATCCCTGCTCTGTTGTTCCGGCTCTTTGTGAGCTTCAAGACCATC 1680 QY
|||||
1621 CCAGTTACCGCCATCCCTGCTCTGTTGTTCCGGCTCTTTGTGAGCTTCAAGACCATC 1680 Db
|||||
1681 CCCACTTACCTGCAATGAGCTCTATCTCTCCATATGTCAGTATGGCTTTGAGGGTGTG 1740 QY
|||||
1681 CCCACTTACCTGCAATGAGCTCTATCTCTCCATATGTCAGTATGGCTTTGAGGGTGTG 1740 Db
|||||
1741 ATCTGAGCATCTATGGCATGAGGAGGACCTGACATGTTTATAGAGAACGCTGCCCG 1800 QY
|||||
1741 ATCTGAGCATCTATGGCATGAGGAGGACCTGACATGTTTATAGAGAACGCTGCCCG 1800 Db
|||||
1801 TTCGGGAGCCACAGAGCATCTCCGAGCGTGGATGTGAGGATGCCAAGCTCTACATG 1860 QY
|||||
1801 TTCGGGAGCCACAGAGCATCTCCGAGCGTGGATGTGAGGATGCCAAGCTCTACATG 1860 Db
|||||
1861 GACTTCTTGGTCTTGGGGATCTTCTTCTAGCCCTGCGGCTGCTGAGGCTACCTTGTGTG 1920 QY
|||||
1861 GACTTCTTGGTCTTGGGGATCTTCTTCTAGCCCTGCGGCTGCTGAGGCTACCTTGTGTG 1920 Db

1861 GACTTCTTGGTCTTGGGCATCTTCTTCTAGCCCTGCGCTGCTGGCCTACCTTGTGTG 1920 Db
|||||
1921 CGTTACCCGGGTCAAGTCAAGAGATAGAGGCTTGGCCAGCCTGTAGCCACCCCTGGA 1980 QY
|||||
1921 CGTTACCCGGGTCAAGTCAAGAGATAGAGGCTTGGCCAGCCTGTAGCCACCCCTGGA 1980 Db
|||||
1981 GCAGGAAGCCCCAGTCCCGAGCCTTTGGGACTGTTTTAACTTATAGACTTTGGCACATG 2040 QY
|||||
1981 GCAGGAAGCCCCAGTCCCGAGCCTTTGGGACTGTTTTAACTTATAGACTTTGGCACATG 2040 Db
|||||
2041 GTTCTTGGGGGGCTATCT 2100 QY
|||||
2041 GTTCTTGGGGGGCTATCT 2100 Db
|||||
2101 GCTTCCAGGCTGGGCTCTGGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT 2160 QY
|||||
2101 GCTTCCAGGCTGGGCTCTGGGAGTGGGGCTCCAGCCCTCCCACTATGCCAGGAGTCT 2160 Db
|||||
2161 TCCCAAGTTGATGCGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220 QY
|||||
2161 TCCCAAGTTGATGCGGTTTGTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220 Db
|||||
2221 CTACTGGGAGGCTGCTGCT 2280 QY
|||||
2221 CTACTGGGAGGCTGCTGCT 2280 Db
|||||
2281 AGCCCTAGGCTCTTAGGGCCCACTTACAACTGACCAAGTGGCCCTCTGGGGGGTCC 2340 QY
|||||
2281 AGCCCTAGGCTCTTAGGGCCCACTTACAACTGACCAAGTGGCCCTCTGGGGGGTCC 2340 Db
|||||
2341 CCACCACACAAGTGTGTTAACTGGGCTGCTATAAGTTGGAGTTTCAGGGCTGGGGCC 2400 QY
|||||
2341 CCACCACACAAGTGTGTTAACTGGGCTGCTATAAGTTGGAGTTTCAGGGCTGGGGCC 2400 Db
|||||
2401 TGGTGGAGTCCACTGGAAGTCCCAATTTATGATGTTGAAATGGACAGGAAGSACTCTGA 2460 QY
|||||
2401 TGGTGGAGTCCACTGGAAGTCCCAATTTATGATGTTGAAATGGACAGGAAGSACTCTGA 2460 Db
|||||
2461 AGTCTTCT 2520 QY
|||||
2461 AGTCTTCT 2520 Db
|||||
2521 GCCAGCAGAGAAGCTGGGTTTTCTGTCTAGGTACCACTTCCCAATCTCTGGGATTTGGA 2580 QY
|||||
2521 GCCAGCAGAGAAGCTGGGTTTTCTGTCTAGGTACCACTTCCCAATCTCTGGGATTTGGA 2580 Db
|||||
2581 GGCCTGGGGCTGTGGGATGCCCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2640 QY
|||||
2581 GGCCTGGGGCTGTGGGATGCCCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2640 Db
|||||
2641 GGTGGCACTGTGCAATAATGTTCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2700 QY
|||||
2641 GGTGGCACTGTGCAATAATGTTCTGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2700 Db
|||||
2701 GCATCTTTATTTGGGGGGGGTGTAGTGGGGAGAACCCCAACCTCTCTCTCTCTCTCTCTCT 2760 QY
|||||
2701 GCATCTTTATTTGGGGGGGGTGTAGTGGGGAGAACCCCAACCTCTCTCTCTCTCTCTCTCT 2760 Db
|||||
2761 CTACGCAATGACGGTCTCTGTTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2820 QY
|||||
2761 CTACGCAATGACGGTCTCTGTTGATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2820 Db
|||||
2821 AGACAGCCATCTACGCCCTACACATGTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880 QY
|||||
2821 AGACAGCCATCTACGCCCTACACATGTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2880 Db
|||||
2881 GGTGCATGCTGTGGTGGCGGGTGTGTTGGGAGGACAGTCCCAACCTCTCTCTCTCTCTCTCT 2940 QY
|||||
2881 GGTGCATGCTGTGGTGGCGGGTGTGTTGGGAGGACAGTCCCAACCTCTCTCTCTCTCTCTCT 2940 Db
|||||
2941 TCCCATGTTGAGACTCTAAGGATTAAGGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000 QY
|||||
2941 TCCCATGTTGAGACTCTAAGGATTAAGGCTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000 Db

```
QY 3001 AGGTGTCTACCCCAAGTCTTCCCTCCTCCCAAGCCAGGGGTGGCACAGGGCACTAGATC 3060
D 3001 AGGTGTCTACCCCAAGTCTTCCCTCCTCCCAAGCCAGGGGTGGCACAGGGCACTAGATC 3060
QY 3061 CTTGGAGTTAGGAACCAACACAAAGCAACACAGGCAACAGGGGATAGTTGGCCCTTGGCCACTGC 3120
D 3061 CTTGGAGTTAGGAACCAACACAAAGCAACACAGGCAACAGGGGATAGTTGGCCCTTGGCCACTGC 3120
QY 3121 CACCCACGGCCCTCCCTTTTGTGTCTCCATGCTGSCATCTTCACTCCCTACCCCTACCCCTCCCA 3180
D 3121 CACCCACGGCCCTCCCTTTTGTGTCTCCATGCTGSCATCTTCACTCCCTACCCCTACCCCTCCCA 3180
QY 3181 GGCACCTGCTCATTCACAACTTCTGTCCATGTCCCTCCACCTGTTCTTATCAGCAGGTGG 3240
D 3181 GGCACCTGCTCATTCACAACTTCTGTCCATGTCCCTCCACCTGTTCTTATCAGCAGGTGG 3240
QY 3241 CCCCTGGGCATCAGACAGCTGCTCCCTGGCCAGCAGGTGGCAGACACTCAGAGCATGT 3300
D 3241 CCCCTGGGCATCAGACAGCTGCTCCCTGGCCAGCAGGTGGCAGACACTCAGAGCATGT 3300
QY 3301 CTGGCTTTCTGCTGGGTCCAGGCTCATCTGCTGCTGATTTCCCTCCCTCCCTCCCTCCCTCA 3360
D 3301 CTGGCTTTCTGCTGGGTCCAGGCTCATCTGCTGCTGATTTCCCTCCCTCCCTCCCTCCCTCA 3360
QY 3361 TTTTCCCTTTTCTGCTGACACATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3420
D 3361 TTTTCCCTTTTCTGCTGACACATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3420
QY 3421 TCCATCACACAGGATGCCAGTTGTTTGTGGG 3455
D 3421 TCCATCACACAGGATGCCAGTTGTTTGTGGG 3455

RESULT 3
US-10-154-452-3
; Sequence 3, Application US/10154452
; Publication No. US20020192821A1
; GENERAL INFORMATION:
; APPLICANT: Roy, Josee
; APPLICANT: Conop, Bruce P.
; TITLE OF INVENTION: INCREASED FUNCTIONAL ACTIVITY AND/OR
; TITLE OF INVENTION: EXPRESSION OF ABC TRANSPORTERS PROTECTS AGAINST THE LOSS OF
; TITLE OF INVENTION: DOPAMINE NEURONS ASSOCIATED WITH PARKINSON'S DISEASE
; FILE REFERENCE: 100103.420
; CURRENT APPLICATION NUMBER: US/10/154,452
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2687
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-452-3

Query Match 75.0%; Score 2592.2; DB 9; Length 2687;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 2619; Conservative 7; Mismatches 15; Indels 2; Gaps 2;

QY 1 GCCACCATGGCGAGAGGGCGTGGAGGGCGGGAACCCCTGCTGCTGACCGACACCTCAAG 120
D 45 GCCACCATGGCGAGAGGGCGTGGAGGGCGGGAACCCCTGCTGCTGACCGACACCTCAAG 120
QY 61 GCCATGGCGTACGCTGGAGGAGGGGGGGAACCCCTGCTGCTGACCGACACCTCAAG 164
D 105 GCCATGGCGTACGCTGGAGGAGGGGGGGAACCCCTGCTGCTGACCGACACCTCAAG 164
QY 121 AAGGTGGAGAACACATCACTGAAGCCCGAGCGCTTCTCCACACTGCCCAAGCGCTCAGCC 180
D 165 AAGGTGGAGAACACATCACTGAAGCCCGAGCGCTTCTCCACACTGCCCAAGCGCTCAGCC 224
QY 181 GTGGACATCGAGTTCTGGAGTGTCTCTATTCGTCGGGAGGGGCGCTGCTGGCGCAA 240
```

```
Db 225 GTGGACATCGAGTTCTGGAGCTGCTCTATTTCGTCGGGAGGGGCGCTGCTGGCGCAA 284
QY 241 AGGGGTATTAAGACCCCTTCTCAAGTGCCTCTCAGGTAAATTTCTGCCGCGGAGCTGATTT 300
D 285 AGGGGTATTAAGACCCCTTCTCAAGTGCCTCTCAGGTAAATTTCTGCCGCGGAGCTGATTT 344
QY 301 GGCATCATGGGCCCCCTCAGGGGCTGCAAGTCTTACATTTATGAACATCTTTGGCAGGATAC 360
D 345 GGCATCATGGGCCCCCTCAGGGGCTGCAAGTCTTACATTTATGAACATCTTTGGCAGGATAC 404
QY 361 AGGAGTCTGGAATGAAGGGGAGATCCTGGTTTAAATGGAAGCCACGGGAGCTTGAGAAC 420
D 405 AGGAGTCTGGAATGAAGGGGAGATCCTGGTTTAAATGGAAGCCACGGGAGCTTGAGAAC 464
QY 421 TTCCGCAAGATGCTCTTACATCATGCAAGATGACATGCTGCTGCCGACCTCAGCGTG 480
D 465 TTCCGCAAGATGCTCTTACATCATGCAAGATGACATGCTGCTGCCGACCTCAGCGTG 524
QY 481 TTGGAAGCCATGATGCTCTGCTAACTTACTGAGATCCCGATGTGAAAC 540
D 525 TTGGAAGCCATGATGCTCTGCTAACTTACTGAGATCCCGATGTGAAAC 584
QY 541 GATCTCGTACAGATCCTGACGGCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
D 585 GATCTCGTACAGATCCTGACGGCCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 644
QY 601 GCCCTGCTCTCTGCGGGGAGAGAGGCTGCTGCCATCCCTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
D 645 GCCCTGCTCTCTGCGGGGAGAGAGGCTGCTGCCATCCCTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 704
QY 661 CGCCCTGCTCATGTTCTTTGATGAGCCACAGTGGTCTGGATAGCGGCTCTGTTTCCAA 720
D 705 CGCCCTGCTCATGTTCTTTGATGAGCCACAGTGGTCTGGATAGCGGCTCTGTTTCCAA 764
QY 721 GTGCTGCTCCTCATGAAGTCCCTGGCACAGGGGGCGGTACCATCATCTGCACCATCCAC 780
D 765 GTGCTGCTCCTCATGAAGTCCCTGGCACAGGGGGCGGTACCATCATCTGCACCATCCAC 824
QY 781 CAGCCAGTGCACCAAGCTCTTTGAGATCTTTGACAAGCTCTACATCTCTGAGCCAGGGTCAG 840
D 825 CAGCCAGTGCACCAAGCTCTTTGAGATCTTTGACAAGCTCTACATCTCTGAGCCAGGGTCAG 884
QY 841 TGCATCTTCAAAGAGTGTGTCACCACTGATCCCTTATCTAAAGGAGCTCGGCTTCAT 900
D 885 TGCATCTTCAAAGAGTGTGTCACCACTGATCCCTTATCTAAAGGAGCTCGGCTTCAT 944
QY 901 TGCCCCACCTTACCACAAACCCGGCTGACTTCATCATGAGTGGCTCTGCGGAGTATGGA 960
D 945 TGCCCCACCTTACCACAAACCCGGCTGACTTCATCATGAGTGGCTCTGCGGAGTATGGA 1004
QY 961 GACCTGAACCCCATGTTTTCAGGGCTGTGCAAGATGGGCTGTGCGGTATGGCTGAGAAG 1020
D 1005 GACCTGAACCCCATGTTTTCAGGGCTGTGCAAGATGGGCTGTGCGGTATGGCTGAGAAG 1064
QY 1021 AAGAGCAGCCCTGAGAAGAACAGGTCCTTCCCTGCCCCATGCCCCCTCTTGTCTCCCGAAGTG 1080
D 1065 AAGAGCAGCCCTGAGAAGAACAGGTCCTTCCCTGCCCCATGCCCCCTCTTGTCTCCCGAAGTG 1124
QY 1081 GATCCCATTTAAAGCCACACCTTTGGCCACAGCACCTTCACAGTTTCTGCATCTCTTTC 1140
D 1125 GATCCCATTTAAAGCCACACCTTTGGCCACAGCACCTTCACAGTTTCTGCATCTCTTTC 1184
QY 1141 AAGAGCAGCTTCTGCTCATCTCAGGACACGCTCTGACCCACCTTACCGTTTCATCTCC 1200
D 1185 AAGAGCAGCTTCTGCTCATCTCAGGACACGCTCTGACCCACCTTACCGTTTCATCTCC 1244
QY 1201 CAGGTGTTATTGGCGTGTCTATCGGCTCTCTTACCTGATATTTGGCGACGATGCCAGC 1260
D 1245 CAGGTGTTATTGGCGTGTCTATCGGCTCTCTTACCTGATATTTGGCGACGATGCCAGC 1304
QY 1261 AAGGTCTTCAACAAACCGGCTGCTCTTCTTCTCCATGCTGTTCTCTCATGTTTCGCGGC 1320
```


Db 2625 GRGCGCTGGGGTGTGGGRTGSOCCATCCCTCCCATCACCTTTGGTGGGGSAGGC 2684
QY 2639 CTG 2641
Db 2685 CTG 2687

RESULT 5

US-10-154-452-7
: Sequence 7, Application US/10154452
: Publication No. US20020192821A1
: GENERAL INFORMATION:
: APPLICANT: Reiner, Peter B.
: APPLICANT: Roy, Josée
: APPLICANT: Connop, Bruce P.
: TITLE OF INVENTION: INCREASED FUNCTIONAL ACTIVITY AND/OR
: TITLE OF INVENTION: EXPRESSION OF ABC TRANSPORTERS PROTECTS AGAINST THE LOSS OF
: TITLE OF INVENTION: DOPAMINE NEURONS ASSOCIATED WITH PARKINSON'S DISEASE
: FILE REFERENCE: 100103.420
: CURRENT APPLICATION NUMBER: US/10/154.452
: NUMBER OF SEQ ID NOS: 9
: CURRENT FILING DATE: 2002-05-22
: SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
: LENGTH: 2687
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-154-452-7

Query Match 74.9%; Score 2589; DB 9; Length 2687;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 2617; Conservative 7; Mismatches 17; Indels 2; Gaps 2;

QY 1 GCCACCATGCGGAGAGGCGCTGGAGCGGTGGGCTGTGGACTAGGGCCGGGGCTGTG 60
Db 45 GCCACCATGCGGAGAGGCGCTGGAGCGGTGGGCTGTGGACTAGGGCCGGGGCTGTG 104
QY 61 GCCATGGCCGTGACGTGGAGGACGGGGCGGAACCCCTGTGTGACACGACACCTGAAG 120
Db 105 GCCATGGCCGTGACGTGGAGGACGGGGCGGAACCCCTGTGTGACACGACACCTGAAG 164
QY 121 AAGGTGGAGAACACATCACTGAAGCCAGCGCTTCTCCACCTGCCCAAGCCCTCAGCC 180
Db 165 AAGGTGGAGAACACATCACTGAAGCCAGCGCTTCTCCACCTGCCCAAGCCCTCAGCC 224
QY 181 GTGACATCGAGTTCGTGGAGCTGCTCTATTCCGTGCGGAGGCGCCCTGCTGCGCCAAA 240
Db 225 GTGACATCGAGTTCGTGGAGCTGCTCTATTCCGTGCGGAGGCGCCCTGCTGCGCCAAA 284
QY 241 AGGGGTTAAGACCCCTTCTCAAGTGCCTCTCAGGTAAATTTCTGCGCGCGGGAGCTGATT 300
Db 285 AGGGGTTAAGACCCCTTCTCAAGTGCCTCTCAGGTAAATTTCTGCGCGCGGGAGCTGATT 344
QY 301 GGATCATGGGCCCCCTCAGGGGTGGCAAGTGTACATATGAAATCTTTGGCAGGATAC 360
Db 345 GGATCATGGGCCCCCTCAGGGGTGGCAAGTGTACATATGAAATCTTTGGCAGGATAC 404
QY 361 AGGAGTCTGGAATGAAGGGCAGATCCTGGTTAATGGAAGGCCACGGAGCTGAGGACC 420
Db 405 AGGAGTCTGGAATGAAGGGCAGATCCTGGTTAATGGAAGGCCACGGAGCTGAGGACC 464
QY 421 TTCGGCAAGATGTCCTGTACATCATCAAGATGACATGTGTGCGCGACCTTCACGGTG 480
Db 465 TTCGGCAAGATGTCCTGTACATCATCAAGATGACATGTGTGCGCGACCTTCACGGTG 524
QY 481 TTGGAAGCCATGATGTCTGTCTAACCCTGAATCTTACTGAGAATCCCGATGGAAGAAC 540
Db 525 TTGGAAGCCATGATGTCTGTCTAACCCTGAATCTTACTGAGAATCCCGATGGAAGAAC 584
QY 541 GATCTCGTGACAGATCTGAGGCGACCTGGGCTCATGTGCTGCCACACGAGGACA 600
Db 585 GAGCTGGTGACAGATCTGAGGCGACCTGGGCTCATGTGCTGCCACACGAGGACA 644

QY 601 GCCTGTCTCTGCGGGCAGAGAAAGCGTCTGCCCATCGCCCTGAGCTGGTCAACAAC 660
Db 645 GCCTGTCTCTGCGGGCAGAGAAAGCGTCTGCCCATCGCCCTGAGCTGGTCAACAAC 704
QY 661 CCGCTGTCTATGTTCTTTGATGAGCCACACAGTGGTCTGGATAGCCCTCTTTTCCAA 720
Db 705 CCGCTGTCTATGTTCTTTAATGAGCCACACAGTGGTCTGGATAGCCCTCTTTTCCAA 764
QY 721 GTGGTCTCCTCATGAAGTCCCTGGCAGACAGGGGGCGGTACCATCATCTGCACCATCCAC 780
Db 765 GTGGTCTCCTCATGAAGTCCCTGGCAGACAGGGGGCGGTACCATCATCTGCACCATCCAC 824
QY 781 CAGCCAGTGCCTCAAGCTCTTTGAGATGTTTGACAAAGCTCTACATCTCTGAGCCAGGCTCAG 840
Db 825 CAGCCAGTGCCTCAAGCTCTTTGAGATGTTTGACAAAGCTCTACATCTCTGAGCCAGGCTCAG 884
QY 841 TGCATCTTCAAGAGGTGTGTCACCAACCTGATCCCTCTATCTAAAGGACTTCGGCTTGCAT 900
Db 885 TGCATCTTCAAGAGGTGTGTCACCAACCTGATCCCTCTATCTAAAGGACTTCGGCTTGCAT 944
QY 901 TGCCCCACCTTACCACCAACCCGGCTGACTTCATCATCGAGGTGGCCTCTGGCGAGTATGGA 960
Db 945 TGCCCCACCTTACCACCAACCCGGCTGACTTCATCATCGAGGTGGCCTCTGGCGAGTATGGA 1004
QY 961 GACCTGAACCCCATGTTGTTGAGGCTGTGCAGAAATGGGCTGTGCGCTATGGCTGAGAAG 1020
Db 1005 GACCTGAACCCCATGTTGTTGAGGCTGTGCAGAAATGGGCTGTGCGCTATGGCTGAGAAG 1064
QY 1021 AAGAGCAGCCCTGTGAGAAGAACAGAGTCCCTGCCCCATGCCCTCTTGTCTCCGGAAGTG 1080
Db 1065 AAGAGCAGCCCTGTGAGAAGAACAGAGTCCCTGCCCCATGCCCTCTTGTCTCCGGAAGTG 1124
QY 1081 GATCCCATTTGAAGCCACACCTTTGGCAGCAGACAGCCCTCACACAGTCTGTCATCTCTTC 1140
Db 1125 GATCCCATTTGAAGCCACACCTTTGGCAGCAGACAGCCCTCACACAGTCTGTCATCTCTTC 1184
QY 1141 AAGAGCAGCTTCTGCTGCTCATCTCAGGGACACAGGCTCTGACCCACCTACGGTTCATGTCC 1200
Db 1185 AAGAGCAGCTTCTGCTGCTCATCTCAGGGACACAGGCTCTGACCCACCTACGGTTCATGTCC 1244
QY 1201 CAGCTGGTTATTTGGCGTGTCTATCGGCTCTCTACCTGATATTGGCGACGATGCGCAGC 1260
Db 1245 CAGCTGGTTATTTGGCGTGTCTATCGGCTCTCTACCTGATATTGGCGACGATGCGCAGC 1304
QY 1261 AAGGTCTTCAACAACACCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320
Db 1305 AAGGTCTTCAACAACACCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1364
QY 1321 CTCATGCCAAGTGTGCTCAGCTTCCCTTTAGAGATGGCGGTCTTCTATGAGGGAGCACCTC 1380
Db 1365 CTCATGCCAAGTGTGCTCAGCTTCCCTTTAGAGATGGCGGTCTTCTATGAGGGAGCACCTC 1424
QY 1381 AACTACTGTGACAGCTCAAGCGGTATTACCTGGCCCAAGACCATGCTGACGTGCGCTTTT 1440
Db 1425 AACTACTGTGACAGCTCAAGCGGTATTACCTGGCCCAAGACCATGCTGACGTGCGCTTTT 1484
QY 1441 CAGGTGCTGTGCGGCTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCT 1500
Db 1485 CAGGTGCTGTGCGGCTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCT 1544
QY 1501 GAGACCAGCGGCTTCTGCTCTTCTCAGCCCTGGCCACCGCCCTTGGTGGGCCAA 1560
Db 1545 GAGACCAGCGGCTTCTGCTCTTCTCAGCCCTGGCCACCGCCCTTGGTGGGCCAA 1604
QY 1561 TCTTTGGGCTGTGTGATGCGAGTGTCTTCAACTCCCTACAGTGGCCACTTTTGTGGGC 1620
Db 1605 TCTTTGGGCTGTGTGATGCGAGTGTCTTCAACTCCCTACAGTGGCCACTTTTGTGGGC 1664
QY 1621 CAGTGTACCGCATCCCTGCT 1680
Db 1665 CAGTGTACCGCATCCCTGCT 1724
QY 1681 CCCACTTACCTGCAATGGAGCT 1740

Db	1725	CCACATTACCTGCANTGGAGCTCCTATCCTCTCTATGTAGGTATGGCTTTGAGGGTG	1784
Qy	1741	ATCCTGACGATCTATGGCATGGAGCGAGAGACCTGACATGTTTAGAGGAACGCTGCCCG	1800
Db	1785	ATCCTGACGATCTATGGCATGGAGCGAGGAGACCTGACATGTTTAGAGGAACGCTGCCCG	1844
Qy	1801	TTCCGGGAGCCACAGAGCATCCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATG	1860
Db	1845	TTCCGGGAGCCACAGAGCATCCTCCGAGCGCTGGATGTGGAGGATGCCAAGCTCTACATG	1904
Qy	1861	GACTTCCTGGCTTTGGGCATCTTCTTCTAGCCCTGGCGTGTGCGCTACCTTGTGCTG	1920
Db	1905	GACTTCCTGGCTTTGGGCATCTTCTTCTAGCCCTGGCGTGTGCGCTACCTTGTGCTG	1964
Qy	1921	CGTTACGGGTCAAGTTCAGAGAGATAGAGCTTGCCTCCAGCGCTGTACCCAGCCCTGCA	1980
Db	1965	CGTTACGGGTCAAGTTCAGAGAGATAGAGCTTGCCTCCAGCGCTGTACCCAGCCCTGCA	2024
Qy	1981	GCAGGAAGCCCCAGTCCACAGCCCTTTGGGACGTGTTTAACTTATAGACTTGGGCACTG	2040
Db	2025	GCAGGAAGCCCCAGTCCACAGCCCTTTGGGACGTGTTTAACTTATAGACTTGGGCACTG	2084
Qy	2041	GTTTCTGGGGGCTATPCTCTCCTCCTTGGCTCTCCACAGGCTGGCTGTCGGACTGC	2100
Db	2085	GTTTCTGGGGGCTATPCTCTCCTCCTTGGCTCTCCACAGGCTGGCTGTCGGACTGC	2144
Qy	2101	GCCTCCAGCCTGGGCTCTGGGAGTGGGGGCTCCAGCCCTCCCACTATGCCAAGGAGTCT	2160
Db	2145	GCCTCCAGCCTGGGCTCTGGGAGTGGGGGCTCCAGCCCTCCCACTATGCCAAGGAGTCT	2204
Qy	2161	TCCCAAGTTGATCGGGTTGTAGCTTCTCCTCCTACTCTCTCCAACACCTGCATGCAAGA	2220
Db	2205	TCCCAAGTTGATCGGGTTGTAGCTTCTCCTCCTACTCTCTCCAACACCTGCATGCAAGA	2264
Qy	2221	CTACTGGAGGCTGTGCTCCTCTCTGCCCCATGGCACTCTCTGCTGTCTGCTGGG	2280
Db	2265	CTACTGGAGGCTGTGCTCCTCTCTGCCCCATGGCACTCTCTGCTGTCTGCTGGG	2324
Qy	2281	AGCCCTAGGCTCTTAGGGCCCCACTTACAACGTAGCAAGTGGCCCCCTCTGGGGTCC	2340
Db	2325	AGCCCTAGGCTCTTAGGGCCCCACTTACAACGTAGCAAGTGGCCCCCTCTGGGGTCC	2384
Qy	2341	CCACACACAAGTGTGTTGTAACATGGGCTGCTATAAGTTGGAGTTCAGGGGCTGGGCC	2400
Db	2385	CCACACACAAGTGTGTTGTAACATGGGCTGCTATAAGTTGGAGTTCAGGGGCTGGGCC	2444
Qy	2401	TGTTGGAGTCCACTGGAAGTCCCATATGGATCTGAAATGGACAGGAGGACTCTGGA	2460
Db	2445	TGTTGGAGTCCACTGGAAGTCCCATATGGATCTGAAATGGACAGGAGGACTCTGGA	2504
Qy	2461	AGTCTCTTCTCCTCTCTCTCTCTCCACCCTAGACCTGGCTGACTTGGACAACTCT	2520
Db	2505	AGTCTCTTCTCCTCTCTCTCTCTCTCCACCCTAGACCTGGCTGACTTGGACAACTCT	2564
Qy	2521	GCCAGGACACAAGCT--GGGTTTTCTGTCTAGTTCACCACTCCCACTCT--GGGATTTGGA	2578
Db	2565	GCCAGGACACAAGCTGGGGTTTTCTGTCTAGTTCACCACTCCCACTCTGGGGGRTTGA	2624
Qy	2579	GAGGCTGGGGCTGTGGGATGCCCACTCCCGCTCCCATCACTCTTGGTGGGGGAGGGC	2638
Db	2625	GRGGCTGGGGGTGGGRTGSCCATCCCGCTCCCATCACTCTTGGTGGGGGAGGGC	2684
Qy	2639	CTG 2641	
Db	2685	CTG 2687	

RESULT 6
US-10-090-455-3
; Sequence 3, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:

: APPLICANT: Chen, Hongyun									
: APPLICANT: Le Bhan, Stephane									
: TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF									
: FILE REFERENCE: 100103.406									
: CURRENT APPLICATION NUMBER: US/10/090.455									
: CURRENT FILING DATE: 2002-03-01									
: NUMBER OF SEQ ID NOS: 17									
: SOFTWARE: FastSeq for Windows Version 4.0									
: SEQ ID NO 3									
: LENGTH: 1941									
: TYPE: DNA									
: ORGANISM: Homo sapiens									
US-10-090-455-3									
Query Match									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1941; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	7	ATGCGGAGAA	CGCTGGAGCG	CGTGGGCTGT	GGACTAGG	CGCGGGGCTGT	GGCCATG	66	
Db	1	ATGCGGAGAA	CGCGCTGGAG	CGCGTGGG	CTGTGGACTAG	CGCGGGGCTGT	GGCCATG	60	
Qy	67	GCGGTGAC	GTGAGGAC	GGGGGGAA	ACCCCTGT	GCTGTGAC	CACGCACTG	126	
Db	61	GCGGTGAC	GTGAGGAC	GGGGGGAA	ACCCCTGT	GCTGTGAC	CACGCACTG	120	
Qy	127	GAGAACCA	CATCACTGA	AGCCCA	CGGCTTCT	CCCACTG	CCCCAAGCC	186	
Db	121	GAGAACCA	CATCACTGA	AGCCCA	CGGCTTCT	CCCACTG	CCCCAAGCC	180	
Qy	187	ATCGAGTT	CGTGAGCT	GTCTTAT	TCCTGCGG	GAGGGCCCTG	CTGGCCAA	246	
Db	181	ATCGAGTT	CGTGAGCT	GTCTTAT	TCCTGCGG	GAGGGCCCTG	CTGGCCAA	240	
Qy	247	TATAAGAC	CCCTTCTCA	AGTGCCCT	CTCAGGTAA	ATTCTTGCCG	CGCGGAGCTG	306	
Db	241	TATAAGAC	CCCTTCTCA	AGTGCCCT	CTCAGGTAA	ATTCTTGCCG	CGCGGAGCTG	300	
Qy	307	ATGGGCCCC	TCAGGGCG	TGGCAAGT	CTACATTT	CATGAACAT	CTTTGGCAGG	366	
Db	301	ATGGGCCCC	TCAGGGCG	TGGCAAGT	CTACATTT	CATGAACAT	CTTTGGCAGG	360	
Qy	367	TCTGGAAT	GAAGGGCA	GATCTCGT	GGTTAA	TGGAAGCC	ACGGAGCTT	426	
Db	361	TCTGGAAT	GAAGGGCA	GATCTCGT	GGTTAA	TGGAAGCC	ACGGAGCTT	420	
Qy	427	AAGATGCT	CTGTACAT	CATGCAAG	ATGACATG	CTGCTGCG	GCACTC	486	
Db	421	AAGATGCT	CTGTACAT	CATGCAAG	ATGACATG	CTGCTGCG	GCACTC	480	
Qy	487	GCCATGAT	GTCTCTGT	CTAACT	CGAATCTTT	ACTGAGA	ATCCGATG	546	
Db	481	GCCATGAT	GTCTCTGT	CTAACT	CGAATCTTT	ACTGAGA	ATCCGATG	540	
Qy	547	GTGACAGA	GATCCTG	ACGGCACT	GGGGCTG	ATGCTGCT	CCACAGG	606	
Db	541	GTGACAGA	GATCCTG	ACGGCACT	GGGGCTG	ATGCTGCT	CCACAGG	600	
Qy	607	CTCTCTG	CGGGGAG	AGGCGT	CTGGCCAT	CGCCCTG	GAGCTT	666	
Db	601	CTCTCTG	CGGGGAG	AGGCGT	CTGGCCAT	CGCCCTG	GAGCTT	660	
Qy	667	GTCATG	PTCTTTG	ATGAGCC	CAAGTGGT	CTGGAT	AGCCCTCT	726	
Db	661	GTCATG	PTCTTTG	ATGAGCC	CAAGTGGT	CTGGAT	AGCCCTCT	720	
Qy	727	TCCTCAT	GAATCCCT	TGGCAC	GGGGCGG	CTACCAT	CTATGC	786	
Db	721	TCCTCAT	GAATCCCT	TGGCAC	GGGGCGG	CTACCAT	CTATGC	780	
Qy	787	AGTGCCA	AGCTCTTT	TGAGATCT	TTTGACA	AGCTCTT	TACATCT	846	
Db	781	AGTGCCA	AGCTCTTT	TGAGATCT	TTTGACA	AGCTCTT	TACATCT	840	

QY 847 TTCAAAGAGTGTGTCACCAACCTGATCCCTATCTAAAGGACTCGCTTGCATTGCCCC 906
Db 841 TTCAAAGAGTGTGTCACCAACCTGATCCCTATCTAAAGGACTCGCTTGCATTGCCCC 900
QY 907 ACCTACCAACACCGGCTGACTTCATCATCGAGGTGGCCTCTCGCGAGTATGAGACCTG 966
Db 901 ACCTACCAACACCGGCTGACTTCATCATCGAGGTGGCCTCTCGCGAGTATGAGACCTG 960
QY 967 AACCCCATGTTGTTACAGGCTGTCAGAAATGGCTGTGCGCTATGCTGAGAGAGAGC 1026
Db 961 AACCCCATGTTGTTACAGGCTGTCAGAAATGGCTGTGCGCTATGCTGAGAGAGAGC 1020
QY 1027 AGCCCTGAGAAGAACAGAGTCCCTGCGCCATGCGCTTGTCTCCGGAAGTGGATCCC 1086
Db 1021 AGCCCTGAGAAGAACAGAGTCCCTGCGCCATGCGCTTGTCTCCGGAAGTGGATCCC 1080
QY 1087 ATTGAAGGCACACCTTTGGCAGCAGACCTCTCACACAGTTCTGCAATCTCTTCAAGAGG 1146
Db 1081 ATTGAAGGCACACCTTTGGCAGCAGACCTCTCACACAGTTCTGCAATCTCTTCAAGAGG 1140
QY 1147 ACCTTCTCTGCAATCTCTAGGAGACGCTCCCTGACCCAGCTACGGTTCATGTCCTCCAGCTG 1206
Db 1141 ACCTTCTCTGCAATCTCTAGGAGACGCTCCCTGACCCAGCTACGGTTCATGTCCTCCAGCTG 1200
QY 1207 GTTATTGGGCTGCTATCGGCTCTCTTCTACCTGCATATTGGCAGCAGTCCGCAAGGTC 1266
Db 1201 GTTATTGGGCTGCTATCGGCTCTCTTCTACCTGCATATTGGCAGCAGTCCGCAAGGTC 1260
QY 1267 TTCAAACACACCGGCTGCTTCTTCTCATGCTGTTCTCTCATGTTTCGCGCGCCCTCATG 1326
Db 1261 TTCAAACACACCGGCTGCTTCTTCTCATGCTGTTCTCTCATGTTTCGCGCGCCCTCATG 1320
QY 1327 CCAACTGTCTACCTTCCCTTAGAGATGGCGTCTTCATGAGCAGCAGCCTCAACTAC 1386
Db 1321 CCAACTGTCTACCTTCCCTTAGAGATGGCGTCTTCATGAGCAGCAGCCTCAACTAC 1380
QY 1387 TGTACAGCCTCAAAAGCGTATTACCTGGCCAGACCATGGCTGACCTGGCCCTTTCAGGTG 1446
Db 1381 TGTACAGCCTCAAAAGCGTATTACCTGGCCAGACCATGGCTGACCTGGCCCTTTCAGGTG 1440
QY 1447 GTGTGTCGGTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCTGAGACC 1506
Db 1441 GTGTGTCGGTGTCTACTGACGATTTGTACTGATGACGGCCAGCCGCTGAGACC 1500
QY 1507 AGCGCTTCTGCTCTCTCTCAGCCCTGCGCACCGCCAGCCGCTTGTGTCGCAATCTTG 1566
Db 1501 AGCGCTTCTGCTCTCTCTCAGCCCTGCGCACCGCCAGCCGCTTGTGTCGCAATCTTG 1560
QY 1567 GGGCTGCTGATCGGAGCTCTTCCAACTCCCTACAGTGGCCACTTTTGTGGGCCAGTT 1626
Db 1561 GGGCTGCTGATCGGAGCTCTTCCAACTCCCTACAGTGGCCACTTTTGTGGGCCAGTT 1620
QY 1627 AGCGCATCCCTGCTCTCTTCTCTCGGCTTCTTTGTGAGTTCAGGTTCAAGACCATCCCTACT 1686
Db 1621 AGCGCATCCCTGCTCTCTTCTCTCGGCTTCTTTGTGAGTTCAGGTTCAAGACCATCCCTACT 1680
QY 1687 TACTGCAATGAGCTCTCTCTCTCTATGTCAGTATGGCTTTGAGGAGTGTGATCCG 1746
Db 1681 TACTGCAATGAGCTCTCTCTCTCTATGTCAGTATGGCTTTGAGGAGTGTGATCCG 1740
QY 1747 AGCATATGATGAGGAGGAGACCTGACATGTTTAGAGGAGGCTGCCCGTTCGG 1806
Db 1741 AGCATATGATGAGGAGGAGACCTGACATGTTTAGAGGAGGCTGCCCGTTCGG 1800
QY 1807 GAGCCACAGCATCTCTCGAGGCTGGATGTGGAGATGCCAAGCTCTACATGACTTC 1866
Db 1801 GAGCCACAGCATCTCTCGAGGCTGGATGTGGAGATGCCAAGCTCTACATGACTTC 1860
QY 1867 CTGGTCTTGGGATCTTCTCTCTAGCCCTGCGGCTGCTGGCCTACCTTGTGCTGGTTAC 1926
Db 1861 CTGGTCTTGGGATCTTCTCTCTAGCCCTGCGGCTGCTGGCCTACCTTGTGCTGGTTAC 1920

QY 1927 CGGTCAGTCAAGTCAAGAGATAG 1947
Db 1921 CGGTCAGTCAAGTCAAGAGATAG 1941
RESULT 7
US-10-037-270-918
; Sequence 918, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungling
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; FILE REFERENCE: 784CIP2B
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: US/10/037,270
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 918
; LENGTH: 3376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2808)
US-10-037-270-918
Query Match 30.7%; Score 1062.4; DB 9; Length 3376;
Best Local Similarity 81.5%; Pred. No. 1e-298;
Matches 1389; Conservative 0; Mismatches 21; Indels 294; Gaps 2;
QY 227 CCTGCTGGCGCAAAAGGGGTTTAAAGACCTTCTCAAGTGCCTCTCAGGTAATTCGCT 286
Db 38 CCGAGCTCACACAGCTGTTTAAAGACCTTCTCAAGTGCCTCTCAGGTAATTCGCT 97
QY 287 GCCGGAGCTGATTTGGCATCATGGCCCTCAGGGCTGGCAAGTCTACATTCATCAACA 346
Db 98 GCCGGAGCTGATTTGGCATCATGGCCCTCAGGGCTGGCAAGTCTACATTCATCAACA 157
QY 347 TCTTGGCAGGATACAGGGAGTCTGGAATGAAGGGGAGATCTGTTTAAATGAAGGCCAC 406
Db 158 TCTTGGCAGGATACAGGGAGTCTGGAATGAAGGGGAGATCTGTTTAAATGAAGGCCAC 217
QY 407 GGGAGCTGAGGACCTTCCCGAAGATGTCCTGCTACATCATGCAAGATGACATGCTGCT 466
Db 218 GGGAGCTGAGGACCTTCCCGAAGATGTCCTGCTACATCATGCAAGATGACATGCTGCT 277
QY 467 CGCACCTACCGTGTGGAGGATGATGCTCTGCTAACCTGATCTTACTGAGATC 526
Db 278 CGCACCTACCGTGTGGAGGATGATGCTCTGCTAACCTGATGAGCTGAGTGAAGC 337
QY 527 CCGATGTGAAAACGATCTGTCGACAGATCTGTCAGGCACTGGGCGCTGATGCTGCT 586
Db 527 CCGATGTGAAAACGATCTGTCGACAGATCTGTCAGGCACTGGGCGCTGATGCTGCT 586

338	AGGAGGTGAAGAAGGAGGCTGGTGACAGAGATCTCTAGCGGCACGTGGCGCCTGATGTCTGTGCT	397
587	CCACACGAGGACAGAGCCCTGCTCTCTGGCGGGCAGAGAACGCTCTGTGGCCATCGCCCTGG	646
398	CCACACGAGGACAGAGCCCTGCTCTCTGGCGGGCAGAGAACGCTCTGGCCATCGCCCTGG	457
647	AGCTGGTCAACAACCCCGCCTGTCTCATGTCTTTTGATGAGCCACCAAGTGTCTGGATAGCG	706
458	AGCTGGTCAACAACCCCGCCTGTCTCATGTCTTTTGATGAGCCACCAAGTGTCTGGATAGCG	517
707	CCTCTCTTTTCCAAAGTGGTCTCCTCATGAAGTCCCTGGGCACAGGGGGCGGTACCATCA	766
518	CCTCTCTTTTCCAAAGTGGTCTCCTCATGAAGTCCCTGGGCACAGGGGGCGGTACCATCA	577
767	TCTGCACCATCCACACGCCAGTGCACAGCTCTTTTGAGATGTTTGTGACAAGCTCTACATCC	826
578	TCTGCACCATCCACACGCCAGTGCACAGCTCTTTTGAGATGTTTGTGACA	625
827	TGAGCCAGGTCAGTGCATCTTCAAGGAGTGGTCACCAACCTGATCCCTATCTAAAGG	886
626	-----AGTGCATCTTCAAGAGCGTGGTCAACAACCTGATCCCTATCTAAAGG	673
887	GACTTCGGCTTTGCATTTGCCCCACCTACACAAACCCGGCTGACTTCATCATCATCGAGTGGCCCT	946
674	GACTTCGGCTTTGCATTTGCCCCACCTACACAAACCCGGCTGACTTCATCATCATCGAGTGGCCCT	733
947	CTGGCAGTATGGAGAGCTGAACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTGTGCG	1006
734	CTGGCAGTATGGAGAGCTGAACCCCATGTTTTCAGGGCTGTGCAGAAATGGGCTGTGCG	793
1007	CTATGGCTGAGAGAAGAGCAGCCCTGAGAAGAACAGAGTCCCTTGCCCCATGCCCCCTT	1066
794	CTATGGCTGAGAGAAGAGCAGCCCTGAGAAGAACAGAGTCCCTTGCCCCATGCCCCCTT	853
1067	GTCTCTCGGAAGTGGATCCCATTTGAAGCCACACCTTTGCGACACAGCACCCCTCACACAGT	1126
854	GTCTCTCGGAAGTGGATCCCATTTGAAGCCACACCTTTGCGACACAGCACCCCTCACACAGT	913
1127	TCTGCATCCTCTTCAAGAGGACCTTCTGTCCATCCTCAGGGACACAGGTCCTGACCCACC	1186
914	TCTGCATCCTCTTCAAGAGGACCTTCTGTCCATCCTCAGGGACAC	959
1187	TACGGTTTCACTGCCACGTTGGTTATTGGCGTGCTCATCGGCGCTCTACTCGCATATTG	1246
960	-----	959
1247	CGCAGATGCCAGCAAGGCTTTCAACAACACCCGGCTGCCTCTTCTTCTCCATGCTGTTC	1306
960	-----	959
1307	TCATGTTTCGCGGCCCTCATGCCAACTGTGCTCACCTTCCCTTAGAGATGGCGGTCTTCA	1366
960	-----	959
1367	TGAGGGAGCACCTCAACTACTGTGTACAGCCTCAAAGCGTATTACTTGGCCCAAGACCATGG	1426
960	-----	959
1427	CTGACGTGCCCTTTACAGTGGTGTCCGGTGGTCTACTACGACCATTTGTACTCTGATGA	1486
960	-----GGTGGTGTCCGGTGGTCTACTACGACCATTTGTACTCTGATGA	1003
1487	CGGGCCAGCCGCTGAGACACAGCGCTTCTGTCTTCTCAGCCCTGGCCACCGCCACCG	1546
1004	CGGGCCAGCCGCTGAGACACAGCGCTTCTGTCTTCTCAGCCCTGGCCACCGCCACCG	1063
1547	CCTTTGGTGCCCAATCTTTTGGGGCTGCTGATCGGAGTGTCTTCCAACTCCCTACAGGTGG	1606
1064	CCTTTGGTGCCCAATCTTTTGGGGCTGCTGATCGGAGTGTCTTCCAACTCCCTACAGGTGG	1123
1607	CCACTTTTGGGGCCAGTTACCGCCATCCCTGTCTTGTGTCTCCGGCTTCTTTGTCTCA	1666
1124	CCACTTTTGGGGCCAGTTACCGCCATCCCTGTCTTGTGTCTCCGGCTTCTTTGTCTCA	1183

RESULT 8

```

US-09-954-531-591
; Sequence 591, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 591
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-591

```

Query Match	28.2%;	Score 972.8;	DB 9;	Length 2930;
Best Local Similarity	70.5%;	Pred No. 1.2e-272;		
Weighted Similarity	51.2;			

Qy	81	GGACGGGGGGAACCCCTGTGCTGACCAACGACCTGAAGAAGGTGGAGAACCATCAC	140
Db	141	GGAGCCACTGAGACGACCTGTGAATGGACATCTGAAAAAAGTAGATAATAACCTCAC	200
Qy	141	TGAAGCCACGGCTTCTCCACCTGCCAAGCGCTCAGCCGTGACACCTGAGATTTCGTGGA	200
Db	201	GGAAGCCACGGCTTCTCCTCTTCTCGGAGGGCAGCTCTGAACATTTGAATTCAGGA	260
Qy	201	GCTCTCTATTCTCGGGAGGGCCCTGCTGCGCAAAAGGGTATATAGACCCCTCT	260
Db	261	CTTTCTCTATTCGGTTCTGAAGACCTGCTGGAGGAAGAGATACAAGACCCCTCT	320
Qy	261	CAAGTGGCTCWCAGGTAAATTCTGCGCGGGAGCGTGAATGGCATCATGGGCCCTCAGG	320
Db	321	GAAGAGGAATTTCCGGGAAGTTCAATAGTGTGAGTTGGTGCCCATTTATGGGTCTCTCCGG	380
Qy	321	GGCTGGCAAGTCTACATTCAATGAACATCTTGGCAGGATACAGGAGGTCTTGAATTAAGAGG	380

Db 381 GCGCGGGAAGTCCACGCTGATGAACATCTCTGGCTGGATACAGGAGACGGGCAATGAAGG 440
QY 381 GCAGATCTCTGTTAATGAAGGACACGGAGCTGAGGACCTTCCGCAAGATGTCCTGCTA 440
Db 441 GCGCGTCTCTCATCAACGGCTGCCCCGGGACCTGCGCTTCCGGAAGGTGCTCTGCTA 500
QY 441 CATCATGCAAGATGACATGCTGCTGCGGACCTCAGGGTGTGGAAGCCATGATGCTCTC 500
Db 501 CATCATGCAAGATGACATGCTGCTGCGGACCTCAGGGTGTGGAAGCCATGATGCTCTC 560
QY 501 TGTAACTGATCTTACTGAGAAATCCGATGTAAGAGATCTCGTGACAGAGATCCT 560
Db 561 GGCACATCTGAAGCTTTCAGAGGAAGATGAAGCAGAAGGAAATGCTCAAGGATACT 620
QY 561 GACGGCACTGGGCTGATGCTGCTGCGGACCTCAGGGTGTGGAAGCCATGATGCTCTC 620
Db 621 GACAGGCTGGGCTGCTGCTTTCGCGCAACACCGCGGACCGGGAGCTGTCAGGTGGTCA 680
QY 621 GAGGAAGCGTCTGGCCATCGCCCTGGAGCTGCTCAACAAACCCGCTGTCATGCTTTGA 680
Db 681 GCGCAAGCGCTGGCCATCGCGTGGAGCTGGTGAACAACCCCTCCAGTCATGCTTCGA 740
QY 681 TGAGCCCAACGAGTGGTCTGGATAGCGGCTCTGTTTCCAAAGTGGTGTCTCCATGAGTC 740
Db 741 TGAGCCCAACGAGGCTGGATAGCGGCTCTGTTTCCAAAGTGGTGTCTCCATGAGTC 800
QY 741 CTGGCACACGAGGCGGCTGACCATCATCTGCACATCCACAGCCAGCTGCCAAGCTCT 800
Db 801 GCTCGCTCAAGGGGTGCTCCATCATTTGCAACCTCCACAGCCAGCCCAACCTCTT 860
QY 801 TGAGATGTTTGAAGCTCTACATCTGAGCCAGGCTCAGTGATCTTCAAGAGTGGT 860
Db 861 CGAGCTGTTTCGACGAGCTTACGCTCTGAGTCAAGGACAATGCTGCTACCGGGGAAAGT 920
QY 861 CACCAACCTGATCCCTATCTAAGGAGCTCGGCTTGCATTTGCCCAACCTACCAACCC 920
Db 921 CTGCAATCTGTGCGCATATTTGAGGATTTGGGCTGGAACCTGCCCAACCTACCAACCC 980
QY 921 GGCTGACTCATCATGAGTGGCTCTGCGGAGTATGAGACCTCAACCCCATGTTGT 980
Db 981 AGCAGATTTTGCATGAGGTGTCATCCGCGGAGTACGGTGATCAGAACAGTGGTGGT 1040
QY 981 CAGGGCTGTGCAAGATGAGGCTGTCGCTATGCTGAGAGAGAGAGAGCCCTGAGAAGAA 1040
Db 1041 GAGAGCGGTTGCGGAGGATGCTGCTGACTCAGACCAACAGAGAGACCTCGGGGTGATC 1100
QY 1041 CGAGGTCCTGCCCCATGCTCTGCTCGGAGTGGATCCCATCAAA----- 1093
Db 1101 CGAGGTGAACCTTTTCTTTGGCAGCGCCCTCTGAAGAGGTAAACGACAAACAGATT 1160
QY 1094 -----GCCACACCTTTGCCACCAAGCAC 1115
Db 1161 AAAGGGTTGAGAAAGGACTCTCTGTCATGGAAGCTGCCACAGCTTCTTGCCAGCTG 1220
QY 1116 CTTACACAGTCTGATCTCTTCAAGAGGACCTTCTGTCATCTCCTCAGGACACGCT 1175
Db 1221 CTTACAGGATCTGATCTCTTCAAGAGGACCTTCTCAGCATCATGAGGACCTGCT 1280
QY 1176 CTTGACCCACCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
Db 1281 CTTGACACACCTGCGGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1340
QY 1236 CTTGATATTGCGGAGTGCAGCAAGCTTCTCAACAAACAGGCTGCTGCTGCTGCTGCT 1295
Db 1341 CTTGGGATCGGAAGCAAGCAAGGCTTTCAGCAACTCGGCTTCTCTCTCTCTC 1400
QY 1296 CATGCTGTTCTCATGTTGCGCGGCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1355
Db 1401 CATGCTGTTCTCATGTTGCGCGGCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1460
QY 1356 GCGGCTTCTCATGAGGAGCACTCAACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1415
Db 1461 GCGAGTCTTCTTCGGAACACCTGAACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1520

QY 1416 CAAGACCAGGCTGACCTGCCCTTTTCAGGTGGTGTGCTCGGCTGCTACTGACGATGTT 1475
Db 1521 CAAGACCAGGAGAGCTGCCCTTTTCAGATCATGTTCCAGTGGCTACTGACGATCGT 1580
QY 1476 GTACTGGATGACGGGCGGCGGCTGAGACCGGCTTCTGCTGCTTCTGACGCCCTGGC 1535
Db 1581 GTACTGGATGACGCTCGACGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1640
QY 1536 CACCGCCAGCGCTTGGTGGCCCAATCTTTGGGGCTGCTGATCGGAGCTGCTTCCAACTC 1595
Db 1641 CACCATGACCTTCCCTGGTGGCACAGTCCCTGGGCTGCTGATCGGAGCGGCTTCCACGTC 1700
QY 1596 CTTACAGTGGGCGGCTTGGTGGGCGGCTTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
Db 1701 CTTGAGTGGGCGGCTTGGTGGGCGGCTTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1760
QY 1656 CTTCTTTGTGAGCTTCAAGACCTTCCACCTTACCTGCAATGAGCTGCTTCTCTCTA 1715
Db 1761 GTTCTCGTCACTTCCGACACCTTCCACCTTACCTGCAATGAGCTGCTTCTCTCTA 1820
QY 1716 TGTCAAGTATGCTTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1775
Db 1821 TGTCAAGTATGCTTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1880
QY 1776 GACATGCT---TTAGAGGAAGCTTCCGCTTCCGCGGAGCCACAGAGCATCTCTCGAGCGCT 1832
Db 1881 GCATGCTGACATCGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1940
QY 1833 GGATGAGGAGTGCAGGCTTACATGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1892
Db 1941 GGAGTGGAAATGCGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2000
QY 1893 CTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1948
Db 2001 CTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2056

RESULT 9

US-10-171-581-276
; Sequence 276, Application US/10171581
; Publication No. US2003010426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10171581
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 276
; LENGTH: 2930
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: X91249
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-276

Query Match 28.2%; Score 972.8; DB 9; Length 2930;

Best Local Similarity 70.5%; Pred No 1.2e-272;

Matches 1351; Conservative 0; Mismatches 517; Indels 48; Gaps 2;

QY 81 GGAGGGGCGGACCCCTTGTGCTGACCCAGCAGCTTCAAGAGGTGGAGAACCATCAC 140
Db 141 GGAGGGCACTGAGAGCGGACCTGCTGAATGGACATCTGAAAGAGTAGATAAATACCTAC 200
QY 141 TGAAGCCGAGCGCTTCTCCACCTGCCCCAAGCGCTGACCGCTGGACATCGAGTTGTTGGA 200
Db 201 GGAAGCCGAGCGCTTCTCTCTGCTGCGGAGCGCTGTGAACATTGAATTCAGGGA 260

```
QY 201 GCTGCTCTATTCGTCGGGAGGGGCGCTCTGCGCGCAAAAGGGTTATAGACCCCTTCT 260
  || || || || || || || || || || || || || || || || || || || || || ||
Db 261 CTTTCTCTATTCGTCGTCGAGGACCTGTCGAGGAGAAAGATACAGACCCCTCT 320

QY 261 CAAGTGGCTCTCAGGTAATCTTCGCGCCGGGAGCTGATTTGGCATCATGCGCCCTCAGG 320
  || || || || || || || || || || || || || || || || || || || || || ||
Db 321 GAAAGGAATTTCCGGGAAGTTCAATAGTGGTGAAGTTGGTGCCATTTATGGTCTCTCCGG 380

QY 321 GCGTGGCAAGCTACATTTCAATCAATCAATCTTTGGCAGGATACAGGAGTCTGGAATGAAGG 380
  || || || || || || || || || || || || || || || || || || || || || ||
Db 381 GCGCGGAAGTCCACGCTGATGAACATCTTGGCTGATACAGGAGGAGCGGCAATGAAGG 440

QY 381 GCAGATCTCTGGTTAATGGAAGCCACGCGGAGCTGAGGACCTTCCGCAAGATGCTCTGCTA 440
  || || || || || || || || || || || || || || || || || || || || || ||
Db 441 GCGCGTCTCTCAACAGCGCTGCGCGGAGCTGCGCTGCTTCCGGAAGTGTCTCTGCTA 500

QY 441 CATCATGCAAGATGACATGCTGCTGCGGACCTCAAGGTTTGAAGCCATGATGCTCTC 500
  || || || || || || || || || || || || || || || || || || || || || ||
Db 501 CATCATGCAAGATGACATGCTGCTGCGGACCTCAAGGTTTGAAGCCATGATGCTCTC 560

QY 501 TGCTAACTCAATCTTACTCAGAAATCCGATGCTGAAAACGATCTCTGACAGAGATCCT 560
  || || || || || || || || || || || || || || || || || || || || || ||
Db 561 GGCACATCTGAGCTTCAGGAGAGGATGAAGCAGAGGAAATGGTCAAGGAGATACT 620

QY 561 GACGGCACTGGGCTGATGCTGCTCCACAGGAGGACAGCCCTGCTCTCTGCGGGCA 620
  || || || || || || || || || || || || || || || || || || || || || ||
Db 621 GACAGCGCTGGGCTGCTGCTGCGGCAACAGCGGAGCGGAGCGCTGTCAGTGGTCA 680

QY 621 GAGGAAGCTGGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680
  || || || || || || || || || || || || || || || || || || || || || ||
Db 681 GCGCAAGCGCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 740

QY 681 TGAGCCACAGTGGTCTGATGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 740
  || || || || || || || || || || || || || || || || || || || || || ||
Db 741 TGAGCCACAGTGGTCTGATGAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 800

QY 741 CTGTCGACAGGGGGCGTACCATCATCTGACACCATCCACAGCCAGTCCCAAGCTCTT 800
  || || || || || || || || || || || || || || || || || || || || || ||
Db 801 GCTCGCTCAAGGGGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 860

QY 801 TGAGATCTTGAAGCTCTACATCTGACCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 860
  || || || || || || || || || || || || || || || || || || || || || ||
Db 861 CGAGCTCTTGCAGCAGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 920

QY 861 CACCAAGCTGATCCCTATCTAAAGGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 920
  || || || || || || || || || || || || || || || || || || || || || ||
Db 921 CTGCAATCTTGCCATATTTGAGGATTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 980

QY 921 GCGTACTTCATCATGAGTGGGCTCTGCGGAGTATGAGACCTGAAACCCCATTTGTTT 980
  || || || || || || || || || || || || || || || || || || || || || ||
Db 981 AGCAGATTTTGTCTATGAGGTTGCTATCGCGGAGTACGGTATCAGAAACAGTCCGGCTGT 1040

QY 981 CAGGGCTGTCAGAAATGGGCTGTCGCTATGCTGAGAAAGAGAGCCCTGAGAGAA 1040
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1041 GAGAGGGTTTGGGAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1100

QY 1041 CGAGGTCCTGCGCCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1093
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1101 CGAGGTCAACCCCTTTCTTGGCAGCCCTCTGAAAGAGTTGAAGAGTTGAAGAGTT 1160

QY 1094 -----GCCACACCTTTGCCACAGCAC 1115
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1161 AAAGGGTTGAGAAAGGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1220

QY 1116 CCTCACACACTTGCATCTCTCTCAGAGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1175
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1221 CCTCACAGCTTCTGCTGCTCTCTCAGAGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1280

QY 1176 CCTCACACACTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1235
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1281 CCTGACACACTGCGCATCCTGCGACATTTGGATCGGCTCTCTATTTGGCTGCTGCTA 1340
```

```
QY 1236 CTTGTCATATTTGGGAGGATCCGACAGGCTTTTAAACAACACGCGGCTCTCTTCTTCTC 1295
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1341 CTTGGGATCGGGAACGAAACCAAGAGGCTTTGAGCAACTCCGCTTCTCTTCTTCTC 1400

QY 1296 CATGCTGTTCTCTATGTTCCGCGCCCTCATGCAACTGCTGCTCACTTCCCTTTAGAGAT 1355
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1401 CATGCTGTTCTCTATGTTCCGCGCCCTCATGCAACTGCTGCTCACTTCCCTTTAGAGAT 1460

QY 1356 GCGGCTTCTCATGAGGAGCAGCTCAACTACTGCTACAGCTCAAAAGCGTATTACCTGGC 1415
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1461 GCGAGTCTTCTTCGGGAACACCTGACTACTGCTACAGCTGAAGGCTTACTACCTGGC 1520

QY 1416 CARAGCATGGCTGAGCTGCGCTTTTTCAGTGGTGTGCTGCTGCTGCTGCTGCTGCT 1475
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1521 CAAGACCATGGCAGAGCTGCGCTTTTTCAGATCATGTTCCAGTGGCTTACTGACATCGT 1580

QY 1476 GTACTGGATCAGCGGCGCAGCGCTGAGACACCGGCTTCTGCTTCTTCAGCGCTGGC 1535
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1581 GTACTGGATGAGCTGCGACCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1640

QY 1536 CACGCGCACCGCTTGGTGGCCCAATCTTTGGGCTGCTGATCGAGCTGCTTCCAACTC 1595
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1641 CACATGACCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1700

QY 1596 CTTACAGTGGGCGCCTTTTGTGGGCGCAGTTACCGCATCCCTGCTGCTGCTGCTGCT 1655
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1701 CTTGAGTGGGCGCCTTTTGTGGGCGCAGTTACCGCATCCCTGCTGCTGCTGCTGCT 1760

QY 1656 CTTTCTTGTGCTGCTTCAAGACCATCCCTTACCTTACCTGCAATGAGCTTCTTCTCTA 1715
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1761 GTTCTTGTGCTGCTTCAAGACCATCCCTTACCTTACCTGCAATGAGCTTCTTCTCTA 1820

QY 1716 TGTGAGTATGGCTTTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1775
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1821 TGTGAGTATGGCTTTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1880

QY 1776 GACATGT---TTAGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1832
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1881 GCACTGTGACATCGACGAGAGCTGCCACTTCCAGAGTCCGAGGAGCTTCTGCGGAGCT 1940

QY 1833 GGATGTGGAGGATGCCAAGCTTACATGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1892
  || || || || || || || || || || || || || || || || || || || || || ||
Db 1941 GGAGGTGGAATTTGCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2000

QY 1893 CTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1948
  || || || || || || || || || || || || || || || || || || || || || ||
Db 2001 CTTCCGCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2056
```

RESULT 10

```
US-10-072-621-5
; Sequence 5, Application US/10072621
; Patent No. US20020169137A1
; GENERAL INFORMATION:
; APPLICANT: Reiner, Peter B.
; APPLICANT: Connop, Bruce P.
; APPLICANT: Pollard, Michelle
; TITLE OF INVENTION: REGULATION OF AMYLOID PRECURSOR PROTEIN EXPRESSION
; TITLE OF INVENTION: BY MODIFICATION OF ABC TRANSPORTER EXPRESSION OR ACTIVITY
; FILE REFERENCE: 100103.402
; CURRENT APPLICATION NUMBER: US/10/072.621
; CURRENT FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-621-5
```

```
Query Match 28.2%; Score 972.8; DB 9; Length 3201;
Best Local Similarity 70.5%; Pred. No. 1.2e-272;
Matches 1351; Conservative 0; Mismatches 517; Indels 48; Gaps 2;
```


Qy	81	GGAGGGGGCGGAACCCCTGTGCTGACCCAGCGACCTGAAGAAGTGGAGAAACACATCAC	140
Db	380	GGAGGCCACTGAGACGACCGCTGTGAATGAGACATCTGAAAAAAGTAGATAATACCTCAC	439
Qy	141	TGAAGCCACGGCTTCTCCACCTGCCCAGCGCTAGCCGTCAGCATCAGTTTCGTGGA	200
Db	440	GGAAGCCACGGCTTCTTCCCTTGGCTCGAGGGGAGCTGTGAACATGAATTCAGGGA	499
Qy	201	GCTGTCTATTTCGGTGGGAGGGCCCTGCTGGCGCAAAAGGGTTATTAAGACCCCTTCT	260
Db	500	CTTTCTCTATTTCGGTTCCTGAAGACCCCTGCTGAGGAGAAAGGATACAGACCCCTCT	559
Qy	261	CAAGTGCTCTCAGGTAAATTTCTGCCGCCGGGAGCTGATTGGCATCATGGGCCCTCAGG	320
Db	560	GAAAGGAATTTCCGGGAAGTTCAATAGTGTGTAGTTGGTGGCCATTATGGTCTCTCCGG	619
Qy	321	GGCTGGCAAGCTTACATTTATGAACATCTTTGGCAGGATACAGGAGTCTCGAATGAAGGG	380
Db	620	GGCCGGGAAGTCCACGCTGATGAACATCTTGGCTGGATACAGGAGACGGGCATGAAGGG	679
Qy	381	GCAGATCCTGGTTTAATGGAAGGCCACGGGAGCTGAGGACCTTCGCGAAGATGCTCTGCTA	440
Db	680	GGCCGTCTCTCAACGGCCTGCCCGGGACCTGCGCTGCTTCCGGAAGGTGTCTCTGCTA	739
Qy	441	CATCATGCAAGATGACATGCTGCTGCCGACCTTCACGGTGTGGAAGCCATGATGGTCTC	500
Db	740	CATCATGCAAGATGACATGCTGCTGCCGACCTTCACGTGTGACGAGAGGCCATGATGGTGT	799
Qy	501	TGCTAACTCAATCTTTACTGGAATCCGATGTGAAAAACGATCTCGTGACAGAGATCCT	560
Db	800	GGCACATCTGAAGCTTCAGGAGAAGGATGAAGCGACAGAGGAAATGGTCAAGAGATACT	859
Qy	561	GACGGCACTGGGCTCATGTCTGCTCCACAGAGACAGCCCTGCTCTCTGGCGGGCA	620
Db	860	GACAGCGCTGGGCTTGTCTTTGCGGCAACACGCGAGCGGAGCGCTGCAGGTGGTCA	919
Qy	621	GAGNACGCTTGGCATCGCCCTGGAGCTGGTTCACAACCCCGCTGTCTGATCTTTTGA	680
Db	920	GGCAAGCGCTGGCCATCGCGCTGGAGCTGGTGTGAACAACCCCTCAGTCAATGCTTCGA	979
Qy	681	TGAGCCCAACAGTGGTCTGGATAGCGCCTTTGTTTCCAAAGTGGTGTCCCTCATGAAGTC	740
Db	980	TGAGCCCAACAGCGGCTGGACAGCGCCTCTCTCTCCAGTGGTCTGCTGATGAAGG	1039
Qy	741	CCTGGCACAGGGGGCGGTACCATATCTGCACCATCCACCGCCAGTCCCAAGCTCTT	800
Db	1040	GCTCGCTCAAGGGGTGGCTCCCATATTTGCACCATCCACAGCCCGCCCAAACTCTT	1099
Qy	801	TGAGATGTTTGAAGCTCTACATCTGAGCCAGGGTCACTGTCATCTTCAAGAGAGTGT	860
Db	1100	CGAGCTGTCGACCACTTTACGTCTGAGTCAAGGACATGTGTACCGGGGAAAAGT	1159
Qy	861	CACCAACCTGATCCCTCATCTAAAGGAGCTCGGCTTGCATTTGCCCCCACTTACCACAACC	920
Db	1160	CTGCAATCTTGTGCCATATTTGAGGGATTTGGGCTTGAATGCCCCAACCTTACCACAACC	1219
Qy	921	GGCTGACTTCATCATCGAGGTGGCTCTGCGCAGTATGGAGACCTGAACCCCATGTTGT	980
Db	1220	ACGAGATTTTGTATCGAGGTTCGATCCGCGGAGTACGGTGTATCAGAAGTACGCGGTGT	1279
Qy	981	CAGGGCTGTGCAATATGGGCTGTGCCTATGGCTGAGAAGAAGAGAGCCCTCAGAGAAG	1040
Db	1280	GAGAGCGGTTGCGGAGGGCATGTGTACTCATGACACACAGAGAGACCTCGGGGGTGATGC	1339
Qy	1041	CGAGGTCCCTTGCCCCATGCCCTCTTGTCTCCGGAAGTGGATCCCATTTGAAA-----	1093
Db	1340	CGAGGTGAACCCCTTTCTTTTGGCACCGGCCCTCTGAAGAGGTTAAAGCAGACAAAAGATT	1399
Qy	1094	-----GCCACACCTTTTCCACACAGCAC	1115
Db	1400	AAAGGGGTTGAGAAAGGACTCTCTGTTCCATGGAAGGCTGCCACAGCTTCTCTCCCACTG	1459

Qy	1116	CCTCACACAGTTCTGCAATCTCTTTCAAGAGGACCTTCTGTGTCCATCCTCAGGAGACACGGT	1175
Db	1460	CCTCACAGCAATTTCTGATCTCTTTCAAGAGGACCTTCTGTGTCCATCCTCAGGAGACACGGT	1519
Qy	1176	CCTGACCCACCTACGGTTTCATGTCCTCCACGCTGGTTATTGGCTGCTCATCGGCTCCTCTCA	1235
Db	1520	CCTGACACACCTCGGCATCACTTCGCACATGGGATCGGCCTCCTCATTTGGCTCGCTGTA	1579
Qy	1236	CCTGCATATTGGGAGGATGCCACGAGGCTTTCAACAACACCGGCTGCTCTTCTTCTC	1295
Db	1580	CTTGGGATCGGAACAGCAAGCAAGAGGCTTTGAGCAACTCCGGCTTCTCTTCTTCTC	1639
Qy	1296	CATGCTGTTCCTCATGTTGGCGGCCCTCATGCCAACTGTGCTCACTTCCCTTTAGAGAT	1355
Db	1640	CATGCTGTTCCTCATGTTGGCGGCCCTCATGCCCTACTGTCTGACATTTCCCTGGAGAT	1699
Qy	1356	GGCGGTCTTCTCAGGAGGACCACTCAACTACTGCTACAGCTCAAAAGCGTATTACCTGGC	1415
Db	1700	GGAGCTTTCTTTCGGGACACCTGAACCTACTGGTACAGCTGAAGGCTACTACCTGGC	1759
Qy	1416	CAAGACCATGGCTGACGTGCCCTTTTCAGGTGGTGTGTCCGGTGGTCTACTGCAGCATTTG	1475
Db	1760	CAAGACCATGGCAGAGCTGCCCTTTTCAGATCATGTTCCAGTGGCTACTTGCAGCATCGT	1819
Qy	1476	GTACTGGATGACGGGCCACGCCGCTGAGACCAGCCGCTTCTGCTCTTCTCAGCCCTGGC	1535
Db	1820	GTACTGGATGACCTCGCAGCCGCTCCGAGCCGCTTGTGCTCTTTGGCGCTGGG	1879
Qy	1536	CACGGCCACCGCCTTGGTGGCCCAACTTTTGGGGCTGCTGATCGAGCTGCTTCCAACTC	1595
Db	1880	CACCATGACCTCCCTGGTGGCAGCTCCCTGGGCTGCTGATCGGAGCCGCTCCACGTC	1939
Qy	1596	CCTACAGTGGCCACTTTTGTGGGCCAGTTACCGCCATCCCTGTCTCTTGTCTCCGG	1655
Db	1940	CCTGCAGTGGCCACTTTTGTGGGCCAGTGAAGCCATCCCGGTGCTCTGCTTCTCCGG	1999
Qy	1656	CTTCTTTGTCAGCTTCAAGACCATCCCACTTACCTGCAATGGAGCTCCTATCTCTCCTA	1715
Db	2000	GTCTTCGTGAGCTTCGACACCATCCCACTACCTACAGTGGATGCTCTACATCTCCTA	2059
Qy	1716	TGTCAGTATGGCTTTTGGGGTGTGATCTCTGACGATCTATGGCATGGAGCGAGGACCT	1775
Db	2060	TGTTCAGTATGGCTTCGAAGGGTTCATCTCTCCATCTATGGCTTATAGACGGGAAGATCT	2119
Qy	1776	GACATGT--TTAGAGGAAGCTGCCGCTTCCGGGAGCCACAGAGCATCCTCCGAGCGCT	1832
Db	2120	GCATGTGACATCGAGAGAGCTGGCCACTTCCAGAAAGTCGGAGGCCATCCTCGGGAGCT	2179
Qy	1833	GGATGTGGAGGATGCCAAGCTCTACATGGACTTCTTGGTCTTTGGGCATCTTCTCTCAGC	1892
Db	2180	GGAGTGGAAAATGCCAAGCTGTACCTGGACTTCATCTACTGCTGGGATTTTCTTCATCTC	2239
Qy	1893	CCTGGCGCTGCTGGCCTACTTGTGTGCTGGTTACCGGGTCAAGTCAGAGAGATAGA	1948
Db	2240	CCTCCGCGCTCATTTGCCATTTTGTCTCAGTACAAATCCGGGACAGAGGTAA	2295

RESULT 11

```

RESOLVER II
US-09-864-761-28360/c
; Sequence 28360, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04

```



```

: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-03-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Anomax Sequence Listing Engine
: SEQ ID NO 28360
: LENGTH: 287
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC000384.1
: OTHER INFORMATION: EXPRESSED IN LUNG, S
: OTHER INFORMATION: EXPRESSED IN BRAIN, S
: OTHER INFORMATION: EXPRESSED IN BONE MAR
: OTHER INFORMATION: EXPRESSED IN ADULT L
: OTHER INFORMATION: EST_HUMAN HIT: AL134
: OTHER INFORMATION: EST_HUMAN HIT: AL134
: OTHER INFORMATION: SWISSPROT HIT: P4584
: US-09-864-761-28360

```

Query Match	6.9%	Score 238;	DB 10;	Length 287;
Best Local Similarity	100.0%;	Pred. No. 4.2e-59;		
Matches 238;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	7	ATGCGGAGAAAGCGCTGGAGGCGTGGGCTGTGGACTAGGCGCGGGGCTGTGCCCATG	66	
Db	238	ATGCGGAGAAAGCGCTGGAGGCGTGGGCTGTGGACTAGGCGCGGGGCTGTGCCCATG	179	
QY	67	GCGGTGACGCTGGAGGACGGGGCGGAACCCCTGTGTCACCGCACCTGAAGAAGGTG	126	
Db	178	GCGGTGACGCTGGAGGACGGGGCGGAACCCCTGTGTCACCGCACCTGAAGAAGGTG	119	
QY	127	GAGAACCAATCACTGAAGCCAGCGCTTCTCCACCTGCCCAAGCGCTAGCGGTGGAC	186	
Db	118	GAGAACCAATCACTGAAGCCAGCGCTTCTCCACCTGCCCAAGCGCTAGCGGTGGAC	59	
QY	187	ATCAGTTCTGGAGCTGTCTATTTCGTGCGGAGGGGCGCTCTGGCGCAAAAGG	244	
Db	58	ATCAGTTCTGGAGCTGTCTATTTCGTGCGGAGGGGCGCTCTGGCGCAAAAGG	1	

RESULT 12

US-09-864-761-11789/c
Sequence 11789, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
SEQ ID NO 11789
LENGTH: 491
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC000384.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.67
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.62
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.67
US-09-864-761-11789

[illegible]

Db 431 GGAACCCCTGTGCTGACCAACGACCTGAAGAAGGTGGAGAACCAACATCACTGAAGCCCA 372
QY 150 GCGCTTCTCCCACTCCCAAGGCTCAGCCGTGGACATCGAGTTCTGAGGAGCTGCTCA 209
Db 371 GCGCTTCTCCCACTCCCAAGGCTCAGCCGTGGACATCGAGTTCTGAGGAGCTGCTCA 312
QY 210 TTCCTGCGGGAGGGGCGCTGTGCGGCCAAAGAGGGT 246
Db 311 TTCCTGCGGGAGGGGCGCTGTGCGGCCAAAGAGGGT 275

RESULT 13
US-09-989-981A-5
; Sequence 5, Application US/0998981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human ABCG5 (hABCG5)
US-09-989-981A-5

Query Match 3.1%; Score 107.2; DB 9; Length 2340;
Best Local Similarity 49.5%; Pred. No. 1.2e-20;
Matches 379; Conservative 0; Mismatches 363; Indels 24; Gaps 3;

QY 212 CCGTGGGAGGGGCGCTGCTGGCCGCAAAAGGGTTATAGACCCTTCTCAAGTCCCTCT 271
Db 264 CCGTGGGAGCATCATCTTCCGGCAGCAGTGGACCGAGATCTCTCAAGATGTCT 323

Db 272 CAGGTAATTTCTCCCGCGGGAGCTGATTGGCATCATGGCCCTCAGGGGTGGCAACT 331
324 CTTGTACGTGGAGAGCGGCGAGATCATGTGATCTTAGAAGCTCAGGCTCCGGGAAA 383

QY 332 CTACATTCATGAACATCTTTGGCAGGATACAGGAGTCTGGAATGA-----AGGGGCAGA 385
Db 384 CCACGCTGTGGAGCGCATGTCCGGAGGCTGGGGCGCGGGGACCTTCTCTGGGGAGG 443

QY 386 TCTCGTTAATGAAGGCCAGCGGAGCTGAGGACCTTCCGCAAGATGCTCCTCATATCA 445
Db 444 TGTATGTGAACGCCGCGGCGCTGCGCGGGAGCAGTTCCAGGACTGCTTCTCCTACGTC 503

QY 446 TCACAGATCACATGCTGTGCGCGACCTCAGCGGTGTTGAAGCATGATGGTCTCTGCTA 505
Db 504 TCGAGGACACACCTCTGTGACAGCTCACCCTGCGCGGAGAGCTGCACTACACGGCG 563

QY 506 ACCTG---AATCTTACTGAGATCCCGATGTGAAACACATCTCGTGACAGAGATCCTGA 562
Db 564 TCTGCGCATCCGCGCGCAATCCGGCTCTCCAGAGAAGGTGGAGCGGCTCATGG 623

QY 563 CGGCACTGGGCTGATGCTGCTGCTCCACACAGACAGACCCCT-----GC 607
Db 624 CAGAGCTGAGTGTAGCCCATGTGGAGACCGGACTGATTGGCAACTACAGCTTGGGGGCA 683

QY 608 TCTCTGGCGGAGAGAGCGCTGCGCCATCGCCCTGGAGCTGGTCAACAACCCGCTG 667
Db 684 TTTCCACGGGTGAGCGGCGCCGGGTCTCCATCGCAGCCAGCGTGTCTCCAGGATCTTAAGG 743

QY 668 TCATGTTCTTTGATGAGCCACAGTGGTCTGGATAGCGCCTCTTGTTCCTCAAGTGGTGT 727
Db 744 TCATGCTGTTGATGAGCCACACAGCGCTGGATGCTGCTAATCAGATTGTCTG 803

QY 728 CCTCATGAAGTCCCTGGCAGAGGGGCGCTGACCATCATCTGACCATCCACAGCCCA,787
Db 804 TCCTCTCTGTGGAACCTGGCTCGCAGGAACCAATGTGTCTCAGCATTCACAGCCCC 863

QY 788 GTGCCAAGCTCTTGGATGTTTGACAGCTCTACATCTGAGCAGGCTGAGTGCATCT 847
Db 864 GTTCTGAGCTTTTTCAGCTCTTTTGACAAAATTTGCCATCTCGGAGAGCTGATTT 923

QY 848 TCAAGGAGTGGTCAACCAACCTGATCCCCCTATCTAAAGGAGCTCGGCTTGCATTGCCCA 907
Db 924 TCTGTGGCAGCGCAGCGGAATGCTTGAATTTCTCAATGACTGCGGTTACCCCTTGTCTG 983

QY 908 CCTACACAAACCGGCTGACTCATCATCATGAGGTGGCCTCTCTGGCGA 953
Db 984 AACATTCAAACCCCTTTTGACTTCTATATGGACCTGACGTCACTGGA 1029

RESULT 14
US-09-837-992-4
; Sequence 4, Application US/09837992
; Patent No. US20020081687A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Schultz, Joshua
; APPLICANT: Shan, Bei
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human sitosterolemia gene (SSG)
; NAME/KEY: CDS
; LOCATION: (107)..(2062)
; OTHER INFORMATION: human sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: protein
US-09-837-992-4

Query Match 3.1%; Score 107.2; DB 10; Length 2340;
Best Local Similarity 49.5%; Pred. No. 1.2e-20;
Matches 379; Conservative 0; Mismatches 363; Indels 24; Gaps 3;

QY 212 CCGTGGGAGGGGCGCTGCTGGCCGCAAAAGGGTTATAGACCCTTCTCAAGTCCCTCT 271
Db 264 CCGTGGGAGCATCATCTTCCGGCAGCAGTGGACCGAGATCTCTCAAGATGTCT 323

QY 272 CAGGTAATTTCTCCCGCGGGAGCTGATTGGCATCATGGCCCTCAGGGGTGGCAACT 331
Db 324 CTTGTACGTGGAGAGCGGCGAGATCATGTGATCTTAGAAGCTCAGGCTCCGGGAAA 383

QY 332 CTACATTCATGAACATCTTTGGCAGGATACAGGAGTCTGGAATGA-----AGGGGCAGA 385
Db 384 CCACGCTGTGGAGCGCATGTCCGGAGGCTGGGGCGCGGGGACCTTCTCTGGGGAGG 443

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 16, 2003, 13:21:33 ; Search time 164 Seconds
(without alignments)
6460.782 Million cell updates/sec

Title: US-10-072-621-4
Perfect score: 3455
Sequence: 1 gccaccatggcggaagc.....atgccagttgtattgtggg 3455
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

ched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: ResSeq_Patents_NA: *
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2.6/ptodata/1/ina/PTCUS_COMB.seq: *
6: /cgn2.6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.2	2.3	2418	4	US-09-245-808-2
2	75	2.2	4403765	4	US-09-103-840A-2
3	75	2.2	4411529	4	US-09-103-840A-1
4	73	2.1	7218	1	US-08-232-463-14
5	69	2.0	6909	4	US-09-199-637A-111
6	63.2	1.8	1518	2	US-08-997-080-88
7	63.2	1.8	1518	2	US-08-997-362-88
8	63.2	1.8	1518	3	US-08-873-970-88
9	63.2	1.8	1518	4	US-09-095-855-88
10	63.2	1.8	1518	4	US-09-324-542-88
11	63.2	1.8	1518	4	US-09-205-426-88
12	62.8	1.8	4403765	4	US-09-103-840A-2
13	62.8	1.8	4411529	4	US-09-103-840A-1
14	61.4	1.8	1926	4	US-09-249-585A-4
15	61.4	1.8	1931	2	US-09-130-114-2
16	59.2	1.7	11672	4	US-08-441-340-2
17	57.8	1.7	20986	4	US-08-961-527-54
18	57.4	1.7	16593	4	US-08-961-527-52
19	57.2	1.7	15611	4	US-09-441-340-1
20	56.8	1.6	13121	4	US-08-961-527-126
21	56.6	1.6	5045	4	US-09-390-721-1
22	56.6	1.6	5045	4	US-09-390-721-3
23	54.6	1.6	13440	4	US-08-961-527-128
24	51.8	1.5	1460	4	US-08-858-207A-49
25	51.8	1.5	8145	4	US-08-961-527-122
26	51.6	1.5	23673	4	US-08-773-816-1
27	50.6	1.5	3489	2	US-08-728-323A-1

c	28	50.6	1.5	3489	4	US-09-298-568-1	Sequence 1, Appli
	29	50.6	1.5	32207	2	US-08-770-379-20	Sequence 20, Appl
	30	50.6	1.5	32207	4	US-08-757-669A-20	Sequence 20, Appl
	31	50.6	1.5	32207	4	US-09-230-371A-20	Sequence 20, Appl
	32	49.2	1.4	732	4	US-08-919-573-3	Sequence 3, Appli
	33	49.2	1.4	735	4	US-08-919-573-1	Sequence 1, Appli
	34	48.8	1.4	28804	2	US-08-592-874-1	Sequence 1, Appli
	35	48.8	1.4	28804	3	US-09-096-942-2	Sequence 2, Appli
	36	48.8	1.4	28804	3	US-09-096-867-2	Sequence 4, Appli
	37	48.2	1.4	1959	4	US-09-061-764A-4	Sequence 17, Appl
	38	48.2	1.4	2061	4	US-09-061-764A-17	Sequence 2, Appli
	39	47.8	1.4	966	2	US-08-766-738-2	Sequence 2, Appli
	40	47.8	1.4	966	4	US-09-262-610-2	Sequence 187, App
	41	47.8	1.4	4635	4	US-09-221-017B-187	Sequence 7, Appli
	42	47.2	1.4	6854	4	US-09-194-905-7	Sequence 48, Appl
	43	47	1.4	25002	4	US-08-961-527-48	Sequence 24, Appl
	44	46.2	1.3	5894	3	US-08-665-259-24	Sequence 24, Appl
	45	46.2	1.3	5894	3	US-08-762-500-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1

US-09-245-808-2
; Sequence 2, Application US/09245808
; Patent No. 6313277
; GENERAL INFORMATION:
; APPLICANT: Doyle, L. Austin
; APPLICANT: Abruzzo, Lynne V.
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which
; TITLE OF INVENTION: encodes it
; FILE REFERENCE: Ross UMB conversion
; CURRENT APPLICATION NUMBER: US/09/245,808
; EARLIER FILING DATE: 1999-02-05
; EARLIER FILING DATE: 1998-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-2

Query Match	2.3%	Score 79.2;	DB 4;	Length 2418;
Best Local Similarity	47.4%	Pred. NO. 2.4e-10;		
Matches 323;	Conservative 0;	Mismatches 338;	Indels 21;	Gaps 2;
QY	295	CTGATTGGCATCATGGGCCCCCTCAGGGGCTGGCAAGTCTACATTCATGAACATCTTTGGCA	354	
Db	461	CTCAAGCCATCTCTGGGACCCAGGTGGAGGCAATCTTCGTTATTAGATGCTTAGCT	520	
QY	355	GGATACAGGAGGCTCTGGAATGAAGGGCAGATCCTGTTAATGAAGGCCAGCGGAGCTG	414	
Db	521	GCAAGGAAAGATCCAAGTGGATTCTGGAGATGTTCTGATAAATGGAGCAGCCGACCT	580	
QY	415	AGGACCTTCGGCAAGATGCTCTCTACATCATCAAGATGACATGCTGCTGCCGACCTC	474	
Db	581	GCGAATTTTAAATGTAATTCAGGTTACGTTAGGTTAGATGTTGTTGATGGGCACTCTG	640	
QY	475	ACGGTGTGGAGCCATGATGGTCTCTGCTGAACCTGAATCTT-----ACTGAGATCCC	528	
Db	641	ACGGTGAAGAAACCTTACAGTTCTCAGAGCTTCTCGGCTTCCACAACTATGACGAAT	700	
QY	529	GATGTGAAAACGATCTCGTGACAGAGATCTTCAGCGCACTGGCGCTGATGCTGCTCC	588	
Db	701	CATGAAAACCAACGAGGATTAACAGGTCATTCAAGAGTTAGTCTCTGGATAAAGTGCA	760	
QY	589	CACACAGGACACCCCTGC-----TCTCTGGCGGAGGAGGAGGAGCTCTG	633	
Db	761	GACTCCAAGTTGGAACACTCAGTTTATCCGTGTTGCTGGAGGAGGAGGAGGAGGACT	820	

	Query Match	2.2%; Score 75; DB 4; Length 4403765;	
	Best Local Similarity	49.0%; Pred. No. 7.6e-08;	
	Matches 263; Conservative	0; Mismatches 265; Indels 9; Gaps 2;	
QY	231	CTGGCCAAAGGGGTTATAAGACCCCTCTCAAGTCGCTCAAGGTAATAATCTGCCGCCG	290
Db	1965431	CTGGACCATCGATGGCGCAAGACACTCTGGACGCATCTCGTTGACGGCGGCCCGG	1965490
QY	291	GGAGCTGATGGCATCATGGGCCCTCTAGGGCTGCCAAGTCTACATTCATGAACATCTT	350
Db	1965491	TATGCTACCGCGCGTCATCGTCTCGTGGCGCGTGCAAGTCGACACTTGTCCCGGTTGGT	1965550
QY	351	GGCAGGATACAGGGA---GTCTGGAATGAAGGGGCAGATCCTCGTTAATGGAAGGCCACG	407
Db	1965551	GGCTGGGTATACGCACCCGACGGATGSCACGGTGACGTTTCGAGGGCCACAACGTTCA	1965610
QY	408	GGAGCTAGGACCTTCGGCAAGATGTCCTGCTACATCATGCAAGATGACATGCTGTGCC	457
Db	1965611	CGAATATGCTCGCTCGGCAGCAGGATCGCGATGTTGCCACAGGACGACGTGGTGACCG	1965670
QY	468	GGACCTACCGGGTGTGGAGCCATGATGGTCTCTGCTAACTCGAATCTTACTCAGAAATCC	527

Db	1974962	GCATCTGCACACAGGGTCCACAAGCTGTCGGGTGTCTACACGACAGCGGGCGCTCGGTGGC	1975021
QY	642	CTTGGAAGCTGGTCAACAACCCGCTGTATGTTCTTTGATAGCCCAACCACTGGTCTGGA	701
Db	1975022	GGTTGAGCTGTTACCGGGCGGCTGCTGCTGATCCTCTGACGACGCCACATATCCGGCGCTAGA	1975081
QY	702	TAGCGGCTCTTGTGTTCCAAAGTGGTGTGCCCTCATGAAGTCCCTGGCAACAGGGGGGCGG	758
Db	1975082	TCTCGCGCTGGACCGCAGAGTATAGACCAATGCTGCGCGCACTGGCGACACCGCGGTG	1975138

RESULT 4
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

	Query Match	2.1%	Score 73;	DB 1;	Length 7218;
	Best Local Similarity	3.2%;	pred. No. 1.4e-08;		
	Matches	10;	Conservative 205;	Mismatches 100;	Indels 0; Gaps 0;
QY	3108	CCTGGCCACTGCCACCGGCCCCTTTTGTGTCATGCTGGCATCTTCACTCCCC	3167		
Db	1134	YVVVYYY	1193		
QY	3168	TACCCTTCGCCAGGCACGCTGCTGCTATTCAAACCTCTGTCCATGTCCCTCCACTGTTC	3227		
Db	1194	YVVVYYY	1253		
QY	3228	TATCAGCAGTGTCCTGGCATCAGACAGCGCTGCCCTGGGCACACAGTGCACACAC	3287		

[illegible]

```

Db 6409 GCGGATCGTCGCGCGCTTGTGTCATGACCCCAAGGTGATGCTGTCGAGCAACCGACCTC 6468
QY 693 TGTCTGGATAGCGCCTCTTGTTCCTCAAGTGGTCTCCCTCATGAACTCCCTGCGCAGGG 752
Db 6469 GCGGCTCGATCCGGAGATGGTGGCGAAGTCTTCGACGTCATGAAGACCCCTGGCCGTGGA 6528
QY 753 GGCCCTACCATCATCTGCACCATCCACGAG 783
Db 6529 AGGCATGACCATGGTCTCGTGACCCACGAG 6559

RESULT 6
US-08-997-080-88
: Sequence 88, Application US/08997080
: Patent No. 5985287
: GENERAL INFORMATION:
: APPLICANT: WATSON, JAMES D.
: APPLICANT: TAN, PAUL L.J.
: TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
: NUMBER OF SEQUENCES: 194
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Law Offices of Ann W. Speckman
: STREET: 2601 Elliott Avenue, Suite 4185
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/997,080
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sleath, Janet
: REGISTRATION NUMBER: 37,007
: REFERENCE/DOCKET NUMBER: 11000.1007
: TELEPHONE: 206-269-0565
: TELEFAX: 206-269-0563
: TELEX:
: INFORMATION FOR SEQ ID NO: 88:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1518 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-997-080-88

Query Match 1.8%; Score 63.2; DB 2; Length 1518;
Best Local Similarity 45.9%; Pred. No. 2.1e-06;
Matches 255; Conservative 0; Mismatches 298; Indels 3; Gaps 1;

QY 281 TCTGCGCGCGGAGCTGATGGCATCATGCGGCCCTCAGGGCTGCGCAAGTCTACATTCGA 340
Db 437 TCGCGCGCGGGAGTCTTCTCCATCTCGCGCCGCTCGGAGTGTGGGAAGACGACACGT 496
QY 341 TGAACATCTTGCAGGATACAGGAGTCTGGATGAAGGGGAGATCCTGGTTAATGAA 400
Db 497 TCGCGATGATCGCGGAGTTCGAGACCCCGGATGAAGGGGCGATCCGCTCGAAGGCGCG 556
QY 401 GGCCACGGGAGCTGAGGACCTTCGCAAGATGCTCTGCTACATCATGTCGAAGATGACATCG 460
Db 557 ACGTGTGAGGAGCCCAACCAACAGCGCAACGTCAACACGGTGTTCAGCACTAGCGCG 616
```

```

QY 461 TGTGTCGCGACCTCACGGTGTGGAAAGCCATGATGCTCTCTGCTAACTGAATCTTACTG 520
Db 617 TGTTCGCGCACATGACGGTCTGGGACACAGTCCGCTACGCGCCGCGCAGCAAGAACTCG 676
QY 521 AGAATCCGATGTGAAGAAACGATCTCGTGACAGAGATCTGAGGGCACTTGGGCTGATGT 580
Db 677 GCAAGGCGGAGGT--CCGCAAGCGCGTGCAGCAGCTGCTGGAGATCGTCCGGCTGACCG 733
QY 581 CGTGTCTCCACAGGAGGACAGCCCTGCTCTGCGGGGCGAGGAGAGCGCTCTGGCCATCG 640
Db 734 AATTCGCGAGCGAGGCGCCCGCCAGCTGTCCGGGCGGAGCAGCAGCGGTGGCTTGG 793
QY 641 CCCTGGAGCTGGTCAACAAACCCGCTGTCTATCTTTTGTATGAGCCCCACAGTGGTCTGG 700
Db 794 CCGGGCACTGGTGAATACCCAGCGCGCTGCTGCTGATGACCGCTCGGAGCGCTCG 853
QY 701 ATAGCGCCTCTTGTTCCTCAAGTGGTGTCCCTCATGAAGTCCCTGGCACAGGGGGCGGA 760
Db 854 ACCTGAAGCTGGCGCCACGTCATGTCAGTTCAGGCTCAAGCGCATCCAGCGGAGTCCGGA 913
QY 761 CCATCATCTGCACCATCCACGCGCCAGTCCCAAGCTCTTTGAGATGTTTGACAGCTCT 820
Db 914 TCAGTTCATCTACGTGACCCAGCAGGACCCAGGAGGCGCTCAGCATGAGTGACCGCATCG 973
QY 821 ACATCCTGAGCCAGG 836
Db 974 CGGTGATGAACCGCG 989

RESULT 7
US-08-997-362-88
: Sequence 88, Application US/08997362
: Patent No. 5985287
: GENERAL INFORMATION:
: APPLICANT: Tan, Paul
: APPLICANT: Hiyama, Jun
: APPLICANT: Visser, Elizabeth
: APPLICANT: Skinner, Margot
: APPLICANT: Scott, Linda
: APPLICANT: Prestidge, Ross
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
: TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
: NUMBER OF SEQUENCES: 194
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Law Offices of Ann W. Speckman
: STREET: 2601 Elliott Avenue, Suite 4185
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/997,362
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
: FILING DATE: June 12, 1997
: APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
: FILING DATE: August 29, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Sleath, Janet
: REGISTRATION NUMBER: 37,007
: REFERENCE/DOCKET NUMBER: 11000.1002c2
: TELEPHONE: 206-269-0565
: TELEFAX: 206-269-0563
: TELEX:
: INFORMATION FOR SEQ ID NO: 88:
```


521	Qy	AGAA	CCCGATGTGAA	AAAGATCTCGT	GACAGAGATCTT	GACGGCACTGGGCC	CTGATGT	580
677	Db	GCAA	AGCGAGGT	---	CCGAAGCGCTCG	ACGAGCTCTGGAGATCT	CGCGTGACCG	733
581	Qy	CGTG	CTCCACAGGAG	ACAGCCCTGCTCT	TGGCGGCAGAGAA	CGCTTGGCCATCG		640
734	Db	AATT	TTCGAGCGC	GAGCGCCG	CAGCTGTCGGCGGC	ACGACGCGGTGG		793
641	Qy	CCCT	GAGCTGGTCA	CAACCCGCTGTC	ATGTTCTTTGATG	ACCCACCACTGCTCG		700
794	Db	CCGG	GCACTGGTGA	ATACCCAGCGCGCTG	CTCGATGAACCGCT	TCGGAGCGCTCG		853
701	Qy	ATAG	CGCTCTTGT	TTCCAGTGTG	TCCCCTCATGAGT	CTCCTGSCACAGGGGCCGTA		760
854	Db	ACCT	GAGCTGGCC	ACGTCATGCAGT	TCGAGCTCAAGCGCAT	CCAGCGGAGGTTCGGGA		913
761	Qy	CCAT	CTATCTGCACC	ATCCAGCCCACTG	TGCCAAGCTCTTTGAGATG	TTTGACAAGCTCT		820
914	Db	TCAG	TTTCATCTAG	TGACCCACGAC	CAGGAAGGGCGCTC	CAGATGAGTGACCCGATCG		973
821	Qy	ACAT	CTGAGCCAGGG					836
974	Db	CGGT	GATGAACG	CGCGG				989

```

RESULT 10
US-09-324-542-88
; Sequence 88, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c1
; CURRENT APPLICATION NUMBER: US/09/324,542
; CURRENT FILING DATE: 1999-06-02
; EARLIER APPLICATION NUMBER: US 08/997,080
; EARLIER FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0

```

QY 581 CGTGCTCCACACGAGGAGACCCCTGCTCTCTGGCGGCGCAGGAGAGCGTCTGGCCATCG 640
Db 734 AATTGGCCGAGCGACGCGCCGCCAGCTGTCCGGCGGCGCAGCAGCGGTGGCTGG 793
QY 641 CCCTGGAGCTGGTCAACAAACCGGCTGTCTCTTTGATGAGCCACCAAGTGTCTGG 700
Db 794 CCGGGCACTGGTGAACATACCCCAAGCGCTGCTCTGATGAACCGCTCGAGCGCTCG 853
QY 701 ATAGCGCTCTGTGTTTCCAAAGTGGTGTCCCTCATGAAGTCCCTGCGCAGGCGGCCGTA 760
Db 854 ACCTGAAGCTGGCCACGCTCATGAGTTCGAGCTCAAGCGCATCCAGCGGAGGTGGGA 913
QY 761 CCATCATCTGCACCATCCACGAGCCAGTGCACCAAGCTCTTTGAGATGTTTGACAAGCTCT 820
Db 914 TCACGTTTCTATCTACGTGACCCAGCAGCAGGAGGCGCTCAGGATGAGTGACCGCATCG 973
QY 821 ACATCCTGAGCGGG 836
Db 974 CGGTGATGAACGCGG 989

RESULT 11
US-09-205-426-88
; Sequence 88, Application US/09205426
; Patent No. 6406704
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Compounds and Methods for Treatment and
; TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
; CURRENT APPLICATION NUMBER: US/09/205,426
; CURRENT FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: 09/095,855
; EARLIER FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 08/997,362
; EARLIER FILING DATE: 1997-12-23
; EARLIER APPLICATION NUMBER: 08/873,970
; EARLIER FILING DATE: 1997-06-12
; EARLIER APPLICATION NUMBER: 08/705,347
; EARLIER FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 88
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Mycobacterium vaccae
09-205-426-88

Query Match 1.8%; Score 63.2; DB 4; Length 1518;
Best Local Similarity 45.9%; Pred. No. 2.1e-06;
Matches 255; Conservative 0; Mismatches 298; Indels 3; Gaps 1;

QY 281 TCTGCCGCGGAGCTGATGTCATATGCGCCCTCAGGGCTCGCAAGTCTACATCA 340
Db 437 TCGCCCGCGGAGTCTTCTCCATGCTCGGCCCTCGCGGTGTGGGAAGACGACCGT 496
QY 341 TGAACATCTGCGAGATACAGGAGTCTGGAATCAAGGCGCAGATCTGGTTAATGGA 400
Db 497 TCGCGATGATCGGGGATTCGAGACCCCGACTGAAGGGCGGATCGCGCTCGAAGCGCG 556
QY 401 GGCACGGGAGCTGAGGACCTTCCGCAAGATGCTCTGCTACATCATGCAAGATGACATGC 460
Db 557 AGCTGTCGAGGACCCACCCACACGACGACGCTCAACAGGTGTTCCAGCACTACGCGC 616
QY 461 TGCTCCGCACTACAGGTGTTGGAAGCATGATGCTCTGCTAACCTGAATCTTACTG 520
Db 617 TGTTCGCGACATGACGCTCTGGGACAACTGCTGCGCTACGCGCCGCGCAGCAAGAACTCG 676
QY 521 AGAATCCGAGTGAACAACTGCTGTCGACAGATCTGTCAGAGATCTGTCAGGCACTGGCGCTGATGT 580
Db 677 GCAAAGCGAGGT-----CCGAGACCGCTCGAGAGCTGCTGGAGATCGTCCGGCTGACCG 733

QY 581 CGTGCTCCACACGAGGAGACCCCTGCTCTCTGGCGGCGCAGGAGAGCGTCTGGCCATCG 640
Db 734 AATTGGCCGAGCGACGCGCCGCCAGCTGTCCGGCGGCGCAGCAGCGGTGGCTGG 793
QY 641 CCCTGGAGCTGGTCAACAAACCGGCTGTCTCTTTGATGAGCCACCAAGTGTCTGG 700
Db 794 CCGGGCACTGGTGAACATACCCCAAGCGCTGCTCTGATGAACCGCTCGAGCGCTCG 853
QY 701 ATAGCGCTCTGTGTTTCCAAAGTGGTGTCCCTCATGAAGTCCCTGCGCAGGCGGCCGTA 760
Db 854 ACCTGAAGCTGGCCACGCTCATGAGTTCGAGCTCAAGCGCATCCAGCGGAGGTGGGA 913
QY 761 CCATCATCTGCACCATCCACGAGCCAGTGCACCAAGCTCTTTGAGATGTTTGACAAGCTCT 820
Db 914 TCACGTTTCTATCTACGTGACCCAGCAGCAGGAGGCGCTCAGGATGAGTGACCGCATCG 973
QY 821 ACATCCTGAGCGGG 836
Db 974 CGGTGATGAACGCGG 989

RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 1.8%; Score 62.8; DB 4; Length 4403765;
Best Local Similarity 54.3%; Pred. No. 8.8e-05;
Matches 127; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 549 GACAGATCTGACGGCACTGGGCTGATGCTGCTCCACACGAGGAGACCCCTGCT 608
Db 3000173 GATGAGCTGGCGCGCTGGCTGGCGCTGCGCGCTAGTGGGCAAGTA 3000114
QY 609 CTCTGGCGGCGAGAGAAAGCTGTGCGCATCCCTGGAGCTGGTCAACAACCCGCTCT 668
Db 3000113 CTCCAAGGCGATGAGATGCGCTGACCTTCCCGAGGTGCTGTATCAACGCCGAGCT 3000054
QY 669 CATGTTCTTTGATGAGCCACCAAGTGTCTGGATAGCGGCTCTTGTTCCTCAAGTGGTGT 728
Db 3000053 GCTGTTCTCGAGAGCCACCTCCCGGCTGGACCGGTGAACGCCCGCAAGATCAAGTA 2999994
QY 729 CCTCATGAAGTCCCTGGCAGAGGGGGGGGTACCATCATCTCCACCATCCACCA 782
Db 2999993 CATCATCTGGACCTGAAGCGCGCGCGCAGCATCTTTCTCACCACGACGA 2999940

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:

[illegible]

```

RESULT 15
US-09-130-114-2
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D9030U1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

```

	Query Match	1.8%	Score 61.4;	DB 2;	Length 1931;
	Best Local Similarity	43.3%;	Pred. No. 6.5e-06;		
	Matches 287;	Conservative 0;	Mismatches 376;	Indels 0;	Gaps 0;
Qy	1029	CCCTGAGAGACGAGGTCCCTGCCCCATCCTCTCTGTCCTCGAAGTGGATCCCAT	1088		
Db	250	CCCTGGTGGCACCTTGTCTCTGTCGTCTCGCCCTCCCGTCTCTGCTCTCCCGT	309		
Qy	1089	TGAAGCCACACCTTTGGCCACCAGCACCCCTCACACAGTTCTGCATCCTCTTCAAGAGGAC	1148		
Db	310	CTCTGTCCTCTCTCCCGGTCTCTGTCCTCTCTCCCGTCTCTCCCGTCTCTCCCGTCTCTCC	369		

```
QY 1149 CTTCTGTCCATCCTCAGGGACAGGCTCTGACCCACCTACGGTTTCATGTCCCACTGGT 1208
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 370 CGTCTCGTCTCTCTCCCGTCCCTCGTCTCTCTCCCGTCTCTCCCGTCTCTCCCGTCTCT 429

QY 1209 TATTGGCGGTGTCTATCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1268
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 430 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 489

QY 1269 CAACAACACCGGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1328
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 490 CCTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 549

QY 1329 AACTGTGCTACCTTCCCTTTAGAGATGGCGGTCTTTCATGAGGGAGCACCTCAACTACTG 1388
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 550 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 609

1389 GTACAGCCTCAAGGGTATTACCTGGGCAAGACCATGGCTACGTGGCCCTTTCAGGTGGT 1448
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
610 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 669

1449 GTGTCGGGTGCTTACTGACATGTGTACTGGATGACGGGCCAGCCCGCTGAGACACAG 1508
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 670 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 729

1509 CCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1568
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 730 CCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 789

1569 GCTGTGATCGGAGGTGCTTCCACTCCCTACAGGTGGCCACTTTTGTGGGCCCCAGTTAC 1628
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 790 CTTCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 849

1629 CGCCATCCCTGTCCTTTGTTCTCGGGCTTCTTTGTACAGTTCAAGACCATCCCACTTA 1688
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 850 CGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 909

QY 1689 CCT 1691
  |||||
Db 910 CCT 912
```

Search completed: June 16, 2003, 13:59:40
Job time : 177 secs


```
Query Match      30.4%; Score 1016.5; DB 22; Length 643;
Best Local Similarity 37.4%; Pred. No. 2.9e-89;
Matches 236; Conservative 103; Mismatches 195; Indels 97; Gaps 11;

QY 52 HLPKRSVDIEFVELSYSVREGPCWKRKYKTLKLSGKFCRRELIGIMGPSGAGKSTF 111
Db 45 YLPSWPAVNIQFSELQYVVDQTNASKT--KTLRRVNGMFHSHeltaITIGPSGAGKTTL 102
QY 112 MNILAGYRESGMGOILVNGRPRELTFKMSYIIMODMLLPHLTVLEAMVYSANLNT 171
Db 103 LNLAGEGAVCEGEILVNNSPDRMVRFRMSRYIMQTDVLDQFVLEMMILAANKLG 162
QY 172 ENPDVKN--DLVTEILTALGLMSCSHTRTALLSGGQRKRRLAIALELVNPNPMFFDEPTS 229
Db 163 KELNLKQKLEVIDEILMLKDLTNTMAQKLSGGERKRLCIALELVNPNPVIFLDEPTT 222
QY 230 GLDSASCFQVSLMKSIAOGRITICTIHOPSAKLEPFEMFDKLYILSQGQCIFKGVVNTLI 289
Db 223 GLDDLSSQCIALLKVLGAGRTVICSIHTPSAKIFEMLDVAVYLAEGQCQVYQCGKSNIV 282
QY 290 PYLKGGLHCHPTVHNPADEFIEVASGEYGD--NPMLFRAVQNGLCAMAEKKSPEKNEV 347
Db 283 PELKNEGLCCPIYNPADFIIEVACLEYGGSYHEPMV-EAVNG-----KV 327
QY 348 PAPCPPCPPPEVDIESHTFATSTLT-----QFCIL 377
Db 328 ARWSPPA-----VEGTTTQTKDTSAGASSFYERQFDEISPKLLEAKCSWMLQYKLL 381
QY 378 FRKTELSILRDTVLTHLRMSHVIVGLIGLLYHIGDGDASKVFNNTGCLFFSMLFMFA 437
Db 382 LVRLMIQWRDKSYIKLFYMNIVLALLVGLSYF-----415
QY 438 ALMPTVLTFPEMAVFMREHLNWSYLSKAYILAKTNADVPFQVVPVYCSIVVMTGQP 497
Db 416 -----VPEINLLKREYNQWYRLSSYYAAMVTSKLPSMFILAVIYLSIVYLMSSQP 467
QY 498 AETSRELLFSALATATALVAQSLGGLIGAASNSLQVATFVGPVTAIPVLFSGFVYSF-- 555
Db 468 LEWFRFAMLTIAFVTAITSDFGLLI--SSRLSLVNAMFMPVLAIVPLILLSIYIGCYGG 526
QY 556 -KTIPTYLOWSSYLSYVRGFEVILTYCMERGD LTC--LEERCPPREPQSILRALDVE 612
Db 527 GTYISPLMRFLMHSYLRHSMESGLVASLYDYGRA DTICEDTEIFCSFKKSKVLLAFIGFE 586
QY 613 DAKLYMDFLVLGIFFLALRLALLVLRVYK 643
Db 587 NMHYLWSLSCLMSFYLLFTVAAYFMIRLRK 617
```

Search completed: June 10, 2003, 16:44:32
Time : 76 secs

QY 337 EKK-----SSP-EKNEVPA-----PCPP----- 353
D 446 -KKVDQLMAAGLITPVTAPVMTSTVPAHEMGOQNTFKPLTPINELSSRVWDSTAGIGNAG 504
QY 354 --CPP-----EVDPIESH-----TFATSLTQFCILFKRTFSLRDTVLTHLR 395
D 505 SCCPKKKKKNSRPAIELDP--SHLCKRONIYATPFYRQSLISLLVRLFWINRDSLTMR 562
QY 396 FMSHVIGVIGLILHIGDDASKVFNNTGCLFFSMLFMAALMPTVLTFFLENAVEMR 455
D 563 FAHLITGLITGLTFYFGINDAAQTLNIFRYLFYIMFYCAFSGILVKFPLEPIYSR 622
QY 456 EHLNWTYSKAYIAKTMADVPQVPCVYCSIVYVMTGQPAETSRFLFSALATATL 515
D 623 EHFENWYSRAYIVAITLADLPQICSALEFVPTVYLTQPLQELWRFCMFFLIVFTAL 682
QY 516 VAQSLGLLIGASNSLOVATFVGPVTAIPVLLFSGFFVSFKTIPTIYLOWSSLSVRYGF 575
D 683 VSQSIGLAVGAAL-SLKLSGILGPPFCFFLQFSGFFLMEKDPVFLRMDFISFLKYSL 741
QY 576 EGVILTIYMERGDLTCLLEERCPPFPREPOSILRALDVEDAKLYMDFLVLGIFFLALRLAY 635
D 742 EGATNAIFGYDREPLACNELYCHLRHPQILASLDWANGNYTLALIFLFGVLVFLRLAF 801
QY 636 LVLRYRVK 643
D 802 YMSFRLR 809

RESULT 10

ABB64566
ID ABB64566 standard; Protein: 609 AA.

XX AC ABB64566;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 20490.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PF 23-MAR-2000; 2000US-191637P.

XX PF 11-JUL-2000; 2000US-0614150.

XX PA (PERE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX PI WPI; 2001-656860/75.

XX DR N-PSDB; ABL08669.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -

XX PS Disclosure; SEQ ID NO 20490; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 609 AA:

Query Match 32.5%; Score 1086.5; DB 22; Length 609;
Best Local Similarity 36.9%; Pred. No. 4.5e-96;
Matches 238; Conservative 118; Mismatches 234; Indels 55; Gaps 10;

QY 1 MAEKALEAVCGCLGPGAVAMAVTLEDGAEPPVTLTHLKKVNHTEAQRSHLPKRSAVD 60
D 1 MADNAVAQAPNGLGP-----QKQKALE 22

QY 61 IFEVELSVSRGPGCMRRKRGYKTLKCLSGKFCRRELIGIMGSPGAGSKTFNNILAGYRE 120

D 23 LHFQSVSYSLKGA---TKGSTPIINEACGVKSGRLTAILGPGSGAGSKTLNALAGFKL 78

QY 121 SGMKGQILVNGRPRELRTFRKMSCYIMQDMLPHLTVLEAMVMSANLNLN---PDVKN 178

D 79 QGVTFQFLNGRPRDIMSFRKMSAYIAQNFVMLNLTVEETLRVSTDLMKPSSTAAQEQ 138

QY 179 DLVTEILFALGMLSCSHRTALLSGQRKRLAIALELVNPNVPMFDEPTSGDLSASQCF 238

D 139 KIIDDIIQLQSCRRTLVKNSLGGHKKRLSIGIELVTNPIMPFDPTSGDLCVGSYQ 198

QY 239 VVSLMKSIAQGGRTIITIHQPSAKLFEMFDKLYLTLSGOCIFKGVVTVNLIPYLKGLGH 298

D 199 VICHQRLAHGRIVVCVVHQPGSRFLQFLDDVVLVLAHGEVLYAGEQRMLPTFAQSGHI 258

QY 299 CPTYHNPADFIIEVASGEYGDNLNPLFRAVQNGLCAMAEKSKSPKNEVPAPCPCPPEV 358

D 259 CPQYNPADFALEVC-SQSTTERCESLITQNM--MHSTASNVMVKLOVDEETALIDVHK 315

QY 359 DPDE-SHTFATSTL---TQFCILFKRTFSLRDTVLTHLRPMHSHVVGVLGLLYHTG 414

D 316 DALDLSHLRGKEQGVFWTQLSVLLRRHLSRMSRDMFAVQMRVLMHVVALLLGVVYVQIG 375

QY 415 DDASKVENNTGCLFFSMLFMAALMPTVLTFFLEMAVEMREHLNAYWSLKAYILAKTWA 474

D 376 GDAQIVSNVSCILFFVILFVFNAGNAMPISLLCMQDSAVFIREYNGVSLGAYYSKVLVA 435

QY 475 DVPFQVCPVYVYCSIVYVMTGQPAETSRFLFSALATATLVAQSLGLLIGASNSLOVA 534

D 436 DLPLQLTCTPTMFISICYFMTGQPPQEPQREAFAMCWLVCVMTAFIGHFGLVIGSLF-TMQLA 494

QY 535 TFGVPTAIPVLLFSGFFVSFKTIPTIYLOWSSYLSYVRYGFGVILTIYGMERGDLTC-- 592

D 495 IFLVPSATIPFLFSGFFIRLNELSWFLRPICDVSFYRIFEGLMRAIYGYDRGECEYA 554

QY 593 LEERCPPFPQSLRALDVEDAKLYMDFLVLGIFFLALRLAYLV 637

D 555 TSNFCYYRTAEQFLKDFQMEGNEFDWMDMAVLGI-FLILLLLAFFV 598

RESULT 11

ABB64724

ID ABB64724 standard; Protein: 615 AA.

XX AC ABB64724;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 20964.

XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX XX

CC The invention relates to an isolated nucleic acid detection reagent.
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 620 AA:

Query Match	37.1%	Score 1241.5	DB 22	Length 620
Best Local Similarity	40.6%	Pred. No. 4.5e-111		
Matches	261	Conservative 103	Mismatches 184	Indels 95
Gaps	11			
46	EAQRFSLPKRSANDIEFVLSYSVRG-	-----PCWRKRGYKTLKLCUSGRF	92	
31	QPKTLQHLPKPAVDLAFHNLTYRVKEGNSHSEGRQWTPPVWTRHNRHNGITGQW- EY	89		
93	CARELIGMGSGACKSTFMNLLAGYRESGMKGQILVNGRPRLRTRFKMSYIWMDDML	152		
	:	:	:	:
90	AAQHFVWL-----QNFEHRRQYDER-----CRAQPKRI-----SOAFC	123		
	:	:	:	:
153	LPHLTVLEAMVMSANLNTENP-----DVKNDL-----VTEILTALGLMSCSHRTALL	201		
124	LHH-----AQOSASRPYPGSGNDCGSEAAQOEIDDLILLFLSEHRYTNTRL	173		
202	SGGQRKRLAIAELVNPNPVMFDEPTSGLDASCSFOVWSLMKSLAQGGRTIICTHOPS	261		
174	SGGQRKRLSIALELVSNPNIMFFDEPTSGLDSSCTFCQIHLKKMLAAGRTVICTHOPS	233		
262	AKLFEMFQKLYILSGOGCIFKGVVNLIPYLKGLGLHCPYHNHPADFIIEVASGEVDLN	321		
234	AKLFEMFQDLYTLADGQCQYOGSTKQLVPPELSTLNCEPSYHNPAISYIEVSGEGHDT	293		
322	PMLFRAVONG-----LCAMAEKSSPEKNVPAACPCCPPEVDPIES	363		
294	RKLDAIDNGKRDVRSSADYAGLKARNDLVKVQNLKAILDKNDA-----S	338		
364	HFFATSTLTCFCLIFKRTFLSILRDTVLTHLRPMSHVVIGVLIGLILHIGDDASKVFN	423		
339	SRYPSTQHQFWVLKRTFLSFSDWDTLMYLRLFAHLVLGFLGILYLYDIGNDGAIVLSN	398		
424	TCCLFFSMLFMFAALMPTVLTFPLEMAVEMRHLNWTSLKAYIYLAKTMAVDFQVCP	483		
399	LGLEFFNMLFMYSMTITILSPLEMPVLKKNFNRYSLKSYLAIISVADLPQAIFC	458		
484	VYCSIVYWMFGQAPETSRFLFSALATATALVAOSLGLLIGAAANSLOVATFVGPVTAI	543		
459	VYIVSIVYIFTSQWELFRFSMFLSACLLISFVAQSVGLVVGAAVN-VONGVFLAPVMSV	517		
544	PVLSSGPFVSKPIPTYLQWSVLSYVRYGFGVILTYGMRGDLTCLERCPRERPO	603		
518	PLLSGFEVSDAIPVYLRWITVLSYIRVGFECTALATYGYREKLCRFQYCHFKSPI	577		
604	STLRALDVEDAKLYMDFLVLVGIFFLALRLAYLVLRYRVKSR	646		
578	TPLEEDMVNANFTLDIALVIVFVLRISAYFLRWKLTVR	620		

RESULT 9

ABB64565
ID ABB64565 standard; Protein; 812 AA.

AC ABB64565:

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 20487.

KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
XX	
XX	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
XX	27-SEP-2001.
XX	
XX	23-MAR-2001; 2001WO-US09231.
XX	
XX	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
XX	(PEKE) PE CORP NY.
PA	
XX	
XX	Venter JC, Adams M, Li PWD, Myers EW;
PI	
XX	WPI; 2001-656860/75.
XX	N-PSDB; ABL08668.
DR	
DR	
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
XX	Disclosure; SEQ ID NO 20487; 2lpp + Sequence Listing; English.
PS	
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 812 AA;
SQ	

Query Match	34.8%	Score	ll63.5;	DB	22;	Length	812;
Best Local Similarity	35.7%;	Pred.	No. 2.4e-103;				
Matches	281;	Conservative	119;	Mismatches	235;	Indels	153;
Gaps	19						
QY	2	A E K A L E A V G C G L G P C A V A M A V T	-----L E D G A E P P V L T T H L K -	-----K V E N H I	----	44	
		: : : : : : : : : :					
D b	29	S E T H L P S R G S G T G S G A L P R P T T P V P S K K L S K R P P H L T L N I T A K S H D A P L A E T N T V L U S T	88				
QY	45	-----T E A O R -----					
		: : : : : : : : : :					
D b	89	S E S Y V G T P L G A T P V S S P A N Q S Q S A O S D S R N T S T A T S G G S I F P H E Q V K T I K K I N I	148				
QY	62	E F V E L S Y V R E G P C W R K R G Y K T L L K L C S G K F C R R E L I G I M G P S G A G K S T E M N I L A G Y R E S	121				
		: : : : : : : : : :					
D b	149	G F E N I R T T F G V Q E R T --K D V I M G L T G Y F K S G E L S A V I G P S G A C T L L N I L S G Y T T Y	206				
QY	122	G M K G Q I L V N G R P R E L R T F R K M S C Y I M Q D M L L P H T L V L E A M M V S A N L N T E N --P D V K N D	179				
		: : : : : : : : : :					
D b	207	G F T G D F R V N G N R R D L K A F K S N V A F I R Q D T S L Q A F L S V K E A M H F A A N L K I G T H M T H S K R E	266				
QY	180	L V T E I L T A L G M S C S H T R T A L L S G G O R K R I A I A L E L V N N P P V M F D E P T S G L D S A S C F Q V	239				
		: : : : : : : : : :					
D b	267	R V R C I L E A I G M Y E N R H T R T G S G G K R L A I A L E L V N N P P V I L D E P T T G L D S E T S N Q L	326				
QY	240	V S L M S L A O G G R T I C T I T H O P S A K L F E M F D K L Y I L S G O C I F K G V V T N I P L Y K L G L H C	299				
		: : : : : : : : : :					
D b	327	I N L L K K A I E G R T V I C T I H O P S A L T F A M F D H L Y A I G E C K I Y A G G A Q N L L F L G A I N L H C	386				
QY	300	P T Y H N P A D F ---I I E V A S G Y G D L N ---P M L F R A V O N G -----I C A M A	336				
		: : : : : : : : : :					
D b	387	P E S Y N P A D Y C E F T W E I A T H D Y D T A E D N Q L E K L V A L M D N G R N E D Y R Q S K T A R V A Q L A A M -	445				

QY 542 AIPVLLFSGFFVSKTIPTIYLOWSSYLVYRGFEGVILTIYGMERDGLTCLLEPCPFRE 601
 DB 585 SVPFLLFSGFFVSNFPAIVLWITVLSYIRYGFECTALATYSFNRTKLQCHAPYCHEFY 644
 QY 602 POSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLRYVKSER 646
 DB 645 PQTLEDMDADFDALDAALCLIFVLLRIFAFLELRWLKSTR 689

RESULT 7

ABB61867

ID ABB61867 standard; Protein; 689 AA.

XX AC ABB61867;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 12393.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX Drosophila melanogaster.

PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-0509231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL05970.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -
 XX PS Disclosure; SEQ ID NO 12393; 21pp + Sequence Listing; English.
 XX CC The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, therapeutics and pharmaceutical drugs. The invention
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL01840-ABL16175) and the encoded proteins
 (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 689 AA;

Query Match 39.5%; Score 1322; DB 22; Length 689;

Best Local Similarity 43.4%; Pred. No. 8.3e-119;

Matches 275; Conservative 126; Mismatches 215; Indels 18; Gaps 10;

QY 22 VTLEDGAEPLPVLTHLKKVENHITAEQRFSLPKRSADVLEFVELSVYREGPCWRK 78

DB 65 VNTNGNEFSKQPLPEVVEEVEHFDYDA--LNNLPAREPDMFEKELSLTVKLG---FN 119

QY 79 RGYKTLKCLSGKRCRRELLIGMPSGAGKSTPMNLAGYRESGMKGQILVNGRPRELRT 138

DB 120 RGSKEILHNVCGRFPGCSQLIAIMGPSGAGKSTLLDLSGFKTTGVDGSLILNGRRDLPS 179

QY 139 FRKMSCYIMQDDMLLPHLVLEAMWVSANLNL--TENPDVKNDLVTEILTALGLMSCSHT 196

DB 180 FRMSCYITQDDRLQPLLTAVNNMHIAADLKLCQTVSBEKESRIEDILLGLYNHDOQT 239
 QY 197 RTALLSGGOKKRLAIALELVNPPVMEFDEPTSGDASCFQVVSIMKSLAOGRTIIC 256
 DB 240 LTMRLSGGOKKRLSIAMELINNPVMEFDEPTTGLDSSCTKVLLELLKLTSGORTIIC 299
 QY 257 IHQPSAKLFEMFEDKLYLSQOGCIFKGVVTNLIPYLKGLGHCPYHNPADFIEVASGE 316
 DB 300 IHQPTAKLFQIFDQVYVLSAGNCVYOGSTQKLVPLQSVLDLPCPMYHNPADYIELACGE 359
 QY 317 YG-DLNPMLFRAVQNGLC-AMAEKKSSPEKNEVPAPCPPEVDPIESHTRF-TSTLTQ 373
 DB 360 YGYDKVDTLTKLATENGSCLTWFHNPASVLAERAEVLMRKYPIPKKT---KSRLEDTSYSNQ 416
 QY 374 FCILPKRTFLSLRLDVTLTHLRFMSHVIGVLIGLLYLGDDASKVENNTGCLFFSMLF 433
 DB 417 CSVLLRRGFIKAKRDTTMTLHRLGVNIAVAALFGAMYDHTREGSVLDONYNLLFAILMH 476
 QY 434 LMFAALMPTVLTFFLEMAVFMREHLNYSKAYYLAKTMADVPFQVGVVVCISIVYWM 493
 DB 477 HSMTTMMLTVLTFPMDISILIKEHFNWYSLKAYYTAMTLVDLPISIISCFEFTVIVYLW 536
 QY 494 TQOPATSRPLLSALATATATVAQSLGLLIGAANSLOVATFVGVPTAIPVLLFSGFFV 553
 DB 537 SYQPMEWIRFFMFFSISLLTVFVGHSGFLMGAWFDVNV-GTFLAPVLTPMMFAGFV 595
 QY 554 SPKTIPTYLOWSSYLVYRGFEGVILTIYGMERDGLTCLLE-ERCPEFPOSTLRALDVE 612
 DB 596 TLRDPSYLRWGSHTSLYRLGEGFISAIYGLDRGLACEEAPYCHYRYPKPFLEETIMR 655
 QY 613 DAKLYMDFLVLGIFFLALRLAYLVLRYVKSER 646
 DB 656 GDOFNNDVIALGVMILVFRFVSIVVLKAKIKSIR 689

RESULT 8

ABB59648

ID ABB59648 standard; Protein; 620 AA.

XX AC ABB59648;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 5736.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-0509231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL03751.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -
 XX PS Disclosure; SEQ ID NO 5736; 21pp + Sequence Listing; English.

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 832 AA;

SO Query Match 51.5%; Score 1723; DB 22; Length 832;

Best Local Similarity 47.1%; Pred. No. 1.6e-157;

Matches 351; Conservative 109; Mismatches 153; Indels 132; Gaps 12;

QY 10 GCGLGCAVAMVLTEDGAPPPVTLTHLKKVENHTEAQR-----PSHLKRSADVIEFV 64

DB 104 GLGNNHALAPKVANNSSGGSP-----NGQKGTIALSHLPORPPVDIEFC 148

QY 65 ELSYSVREGPCWRKRGYKTLKLSGKRCRELIGIMPSCGAGKSTFPMNLAGYREGMK 124

DB 149 DISYSVTD-----HRRGFKTLKSVSGKFRNGEITAIMGPSGAKSTLMNLAGYKTAQLS 205

QY 125 GOILVNGRPRELRTFRKMSYIMODDMLPHLTVLEAMVYSANLNTEN--PDVKNDLV 182

DB 206 GSVLINSKERNLRFRKLSYIMODDVLIANLTVREAMVAAANLKLGNMISYAKVVVE 265

QY 183 EILTALGLSCSTRTALLSGGORKLAIALELVNPPVMEFDEPTSGDASCFQVVS 242

DB 266 EILETGLKESVNTLCNLGGORKLSIALELVNPPVMEFDEPTSGDSSCTFOJLS 325

QY 243 MKSLAOGRTIICITIHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLIPYKLGGLHCPY 302

DB 326 LRSIARGRTIVCTIHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLIPYKLGGLHCPY 385

QY 303 HNPADFTIEVASGEYGDNLPMFLRAVONGLC-----FATSTLTQFCILF 378

DB 386 HNPADYVLEVASGEYGDVAPKLVDAVKGACKKYAHKDYVLTILAQKCNNDIIKSGSGA 445

QY 334 --AWA---EKKSSPEKNEVPAPCCP--PPEVDP----- 360

DB 446 ENAMAILTEDEKPPLEDROLEPSIPVDDPAELKPKLETQQSONSDCVNMPNAVD 505

QY 361 -----IESHT-----FATSTLTQFCILF 378

DB 506 SCFSFSSKQNAVGGSGGSAVVCMTSLDSDSHSVVTLPNKTCGFTSGWTQFWILL 565

QY 379 KRTFLSTLRVTVLTHLRFMSHVIVGLILYHIGDDASKVFNNTGCLFFSMLFMFAA 438

DB 566 KRSEFRTILDKMLTHMLRFSHVTVGAIGMIYYDVGNASKIMSNAGCIPFVSLTFTA 625

QY 439 LMPVLTFFPLEMAVEMREHLNLYSLKAYVLAKTMADVPEQVYCPVYCSIVYMTGPA 498

DB 626 MMTILTFPTFMSVYVREHLNLYSLKAYFAKTIADMPFQIVFSSVYVLYVLYLTSPQM 685

QY 499 ETSRFLFSALATATALVAQSLGILLGAASNSLQVATFVGPTAIPVLLFSGFVSKTI 558

DB 686 ELERSVMFVLICVLSVAGSLGILLGACGN--IETGVFLGPVTITPILFSGFEVNFDTI 744

QY 559 PTYLOWSSLYSYRYGFEVGLTIYMERGDLTCLEERCDFPREPQSILRALDVEDAKLYM 618

DB 745 PGYLOWVTVYVRYGFECAVIAIYMDRAKMCQNMCHYRVPKFKLEMSMDNALFW 804

QY 619 DFIVL-GIFLALRLAYLVLRV 642

DB 805 DAVALIGIFF-AURITAYFVLRWL 828

RESULT 6

AA778980

ID AA778980 standard; Protein: 689 AA.

XX

AC AA778980;

XX

DT 05-JUN-2000 (first entry)

XX

XX

DE Silkorm Bm white 2 amino acid sequence.

XX Bm white 2; silkorm; visual characteristic; eye colour; egg colour.

XX

OS Bombyx mori.

XX JP2000000094-A.

XX

PD 07-JAN-2000.

XX

PF 15-JUN-1998; 98JP-0166944.

XX

PR 15-JUN-1998; 98JP-0166944.

XX

PA (NORQ) NORINSUISANSHO SANSHI KONCHU.

XX

XX WPI: 2000-199543/18.

XX N-PSDB; AA298624.

XX

PT Silkorm Bm white gene - for changing visual characters such as egg

XX colour and eye colour of silkorm

XX Disclosure; Page 11-13; 16pp; Japanese.

XX

CC This sequence represents the silkorm Bm white 2 amino acid sequence. The

CC Bm white 1, 2, and 3 genes are identified in the invention. Bm white 1

CC and 2 are homologous to the white gene of humans and mice, and Bm white 3

CC is homologous to the white gene of the mosquito and fly. The genes can be

CC used for changing visual characteristics such as egg colour and eye

CC colour of the silkorm.

XX

SO Sequence 689 AA;

Query Match 45.2%; Score 1513; DB 21; Length 689;

Best Local Similarity 45.2%; Pred. No. 2.6e-137;

Matches 319; Conservative 97; Mismatches 193; Indels 96; Gaps 8;

QY 13 LGCAVAMAVTLEDGAPPPVTLTHLKKVENHTEAQRFSHLKRSADVIEFVLSYSVRE 72

DB 10 IGKGSVKVTT-----PTQHKTLSHLPKRPQVDFLAFTDLYKVKQK 49

QY 73 GPCWRKRGYKTLKLSGKRCRELIGIMPSCGAGKSTFPMNLAGYRESCKQILVNGR 132

DB 50 G-----RKSNTKILNSVSGRLSGELTAILGPSGAGKSTILNLTGNTSMGSGITVNGM 106

QY 133 PRELRTFRKMSYIMODDMLPHLTVLEAMVYSANLNTENP--DYKNDILVTILTLALGL 190

DB 107 ERNLSFRKLSYIMODDMLPHLTVLEAMVYSANLNTENP--DYKNDILVTILTLALGL 166

QY 191 MSCSHTRTALLSGGORKLAIALELVNPPVMEFDEPTSGDASCFQVYVLSMKSQAQG 250

DB 167 MEHRKTMTSNLGGQKRLSIALELVNPPIMFDEPTSGDSSSCFQCISLLKLTASEG 226

QY 251 RTIICITHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLIPYKLGGLHCPYHNPADFTI 310

DB 227 RTIICITHOPSAKLFEMFDKLYLSQGQCIFKGVVNTLIPYKLGGLHCPYHNPADFTI 286

QY 311 EVASGEYGDNLPMFLRAVONG-----LC 333

DB 287 EVSCEYGVNNTGKLVRAIDNGNNDIRNGMPFPDQTPDYNKKDMEVSLRAGSNWKNLDS 346

QY 334 AMAEKSSPEKN-----EVPAPCCPPEVDPIES----- 363

DB 347 QVOEKFKDNNSGANGHGNLQILOYATSEV-SKGEPLSLQMDTKQDNAAAALLGTEG 405

QY 364 --HTEATSTLTQFCILFRTFLSILRDTVLTHLRFMSHVIVGLILYHIGDDASKYFV 421

DB 406 SPERYATSEWKQFVVLKTLFESRRDWTLMYURLFAHLVGLFGLYDIDGDDGSKVL 465

QY 422 NNTGCLFFSNLFLMFAALMPTVLTFPLEMAVEMREHLNLYSLKAYVLAKTMADVFPQV 481

DB 466 SNLGFELFNNLFLMFTSMTITILSPLEMPVLVKEHFNWYSLRSYLAITYSDIFPQAI 525

QY 482 CPVYCSIVYVMTGQPAETSRFLFSALATATALVAQSLGILLGAASNSLQVATFVGPVT 541

DB 526 FCIIYVIVVILTSQPPVWFRFAMFLSSCILLISFVAQSVGLVVGAAAMN-VQGVFLAPVM 584

KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
 KW immunosuppressive; gene therapy; cytokine cell proliferation;
 KW cell differentiation modulator; immune disorder; infection; cancer;
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.

XX Homo sapiens.

OS WO200153453-A2.

XX 26-JUL-2001.

XX 23-DEC-2000; 2000WO-US34960.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0523117.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0820312.

XX 03-AUG-2000; 2000US-0853450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 30-NOV-2000; 2000US-0250383.

(HYSE-) HYSEQ INC.

PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;

PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Drmanac RT;

DR WPI: 2001-488707/53.

DR N-PSDB: AAH90113.

XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful

PT for treating e.g. cancer and immune deficiency disorders.

XX Claim 10; Page 639-640; 648pp; English.

XX The present sequence is one of 251 novel human polypeptides encoded
 CC by a bone marrow-expressed polynucleotide. The polynucleotide and the
 CC polypeptide encoded by it are useful in the treatment of various
 CC immune deficiencies and disorders. The deficiencies and disorders may
 CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
 CC infection, or may result from an autoimmune disorder, a coagulation
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
 CC suppression of an inflammatory response or treatment of a nervous
 CC system disorder such as Alzheimer's disease. Detection of the presence
 CC or increased expression of the polynucleotide or the protein it
 CC encodes is useful for the diagnosis and/or prognosis of one
 CC or more types of cancer. The polynucleotide and polypeptide can be
 CC used as nutritional sources or supplements and in the screening of
 CC chemical compounds as potential drugs.

SO Sequence 935 AA;

Query Match 68.7%; Score 2299; DB 22; Length 935;

Best Local Similarity 81.3%; Pred. No. 3e-213;

Matches 457; Conservative 3; Mismatches 4; Indels 98; Gaps 2;

QY 80 GYTKLCLSGKRCRELIGIMGPGAGKSTFNILAGYRESGKMGQILVNGRPRELRTF 139

DB 19 GYTKLCLSGKRCRELIGIMGPGAGKSTFNILAGYRESGKMGQILVNGRPRELRTF 78

QY 140 RKMSCIYIMQDDMLLPHLTVLEAMVMSANLMTENPDVKNDLVTILTLGLMSCSHTRTA 199

DB 79 RKMSCIYIMQDDMLLPHLTVLEAMVMSANLMTENPDVKNDLVTILTLGLMSCSHTRTA 138

QY 200 LLSGGQKRKRAIALELVNPNPVMFFDEPTSGLSASCFQVVSMLKSLAQGGRTICTIHQ 259

DB 139 LLSGGQKRKRAIALELVNPNPVMFFDEPTSGLSASCFQVVSMLKSLAQGGRTICTIHQ 198

QY 260 PSKALFEMFDKLYLSGQOCIFKGVVNTLIPYLKGLGLHCPYHNPAFIIIVASGEYGD 319

DB 199 PSKALFEMFDK-----CIFKGVVNTLIPYLKGLGLHCPYHNPAFIIIVASGEYGD 250

QY 320 LNPMLFRAVONGLCAMAEKSSPEKNEVPAPCPPEVDPVIESHTFATSTLTQFCILFK 379
 DB 251 LNPMLFRAVONGLCAMAEKSSPEKNEVPAPCPPEVDPVIESHTFATSTLTQFCILFK 310
 QY 380 RTFSLSLRDVTLTHLRFMSHVVGVLIGLLYLHIGDDASKVFNNTGCLFFSMLFLMAFAL 439
 DB 311 RTFSLSLRDVT----- 320
 QY 440 MPTVLTFFPELVMAVFMREHLNMYSLKAYLAKTMAVDFVQVGVVYCSIVYVMTGCPAE 499
 DB 321 -----WCVPVYCSIVYVMTGCPAE 340
 QY 500 TSRELLFSALATATATVAQSLGLLIGAASNSLOVATFVGPTAIPVLLFSGFFVSEKTIIP 559
 DB 341 TSRELLFSALATATATVAQSLGLLIGAASNSLOVATFVGPTAIPVLLFSGFFVSEKTIIP 400
 QY 560 TYLQWSSYLSYVRYGFEVILTYIGMERGDLTCLERCPCPREPOSTILRALDVEDAKLYMD 619
 DB 401 TYLQWSSYLSYVRYGFEVILTYIGMERGDLTCLERCPCPREPOSTILRALDVEDAKLYMD 460
 QY 620 FLVLGIFFLALRLAYLVLYR 641
 DB 461 FLVLGIFFLALRLAYLVLYR 482

RESULT 5

ABBS59544

ID ABB59544 standard; Protein: 832 AA.

XX ABB59544;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 5424.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB; ABL03647.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

XX Disclosure; SEQ ID NO 5424; 21pp + Sequence Listing; English.

PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABU01840-ABU16175) and the encoded proteins
 CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO

Db 282 ELFDQYLVLSQGCYVRGKSNLVPYLRDLGLNCPTYNHPADFVMEVASGEYGDNSRLV 341
QY 326 RAYONGLCMAEAKSSPEKNEVPAPCPPEVDPIE---SHTFATSTLTQFCILFKRTF 382
Db 342 RAVREGMCDADYKRDLDGDDVAPFLWHRPAEDSASMECHSFASCLTQFCILFKRTF 401
QY 383 LSLTRDTVLTHLRFMSHVIVGLIGLLYLHIGDDASKVFVNTGCLFFSMLFLMFAALMPT 442
Db 402 LSIARDSVLTHLRITSHIGICLLGLLGLYLGIGNEAKKVLNSGFLFFSMLFLMFAALMPT 461
QY 443 VLFPPEMAYFMREHLNLYWYSLKAYLAKTMADVFPQVCPVVCYSIVYVWMTQPAETSR 502
Db 462 VLFPPEMYSVFLREHLNLYWYSLKAYLAKTMADVFPQIMFPVAYCSTVYVWMTQPSDAVR 521
QY 503 FLFLSALATATATVAOSGLLGAANSLOVATFVGPVTAIPVLLRSGFFVSEKTIPTYL 562
Db 522 FLVLAALGTWTSILVAOSGLLGAANSLOVATFVGPVTAIPVLLRSGFFVSEKTIPTYL 581
563 QWSSYLSYVRYGPEGVILTYIGMERGDLTC-LEERCPPREPOQSILRALDVEDAKLYMDFL 621
Db 582 QWSSYLSYVRYGPEGVILTYIGMERGDLTC-LEERCPPREPOQSILRALDVEDAKLYMDFL 641
QY 622 VLGIFFLALRLAYLVLYRYKSER 646
Db 642 VLGIFFISLRLEYVLYRYKIRAER 666

RESULT 3
AAM78589
ID AAM78589 standard; Protein: 935 AA.
AC AAM78589;
XX
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 1251.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WO200157190-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX 20-JUN-2000; 2000US-0598075.
XX 19-JUL-2000; 2000US-0620325.
XX 01-SEP-2000; 2000US-0654936.
XX 15-SEP-2000; 2000US-0663561.
XX 20-OCT-2000; 2000US-0693325.
XX 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX N-PSDB; AAK51722.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX
XX Claim 20: Page 3510-3512; 6221pp; English.

xx The invention relates to polynucleotides (AAK51456-AAK53435) and the
cc encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
cc cytokine, cell proliferation or cell differentiation or which may induce
cc production of other cytokines in other cell populations. The
cc polynucleotides and polypeptides are useful in gene therapy, vaccines or
cc peptide therapy. The polypeptides have various cytokine-like activities,
cc e.g. stem cell growth factor activity, haematopoiesis regulating
cc activity, tissue growth factor activity, immunomodulatory activity and
cc activin/inhibin activity and may be useful in the diagnosis and/or
cc treatment of cancer, leukaemia, nervous system disorders, arthritis and
cc inflammation.
cc Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
cc (AAM80020) are omitted as the relevant pages from the sequence listing
cc were missing at the time of publication.
xx
xx Sequence 935 AA;
SQ
Query Match 68.7%; Score 2299; DB 22; Length 935;
Best Local Similarity 81.3%; Pred. No. 3e-213;
Matches 457; Conservative 3; Mismatches 4; Indels 98; Gaps 2;
QY 80 GYKTLKCLSGKFCRRELIGIMPGSGAGKSTFNILAGYRESGKMGQILVNGRPRELRTF 139
Db 19 GYKTLKCLSGKFCRRELIGIMPGSGAGKSTFNILAGYRESGKMGQILVNGRPRELRTF 78
QY 140 RKMSYIMQDDMLLPHLTVLEAMVSNANLNTENPDVKNLVTETLTALGLMSCSHRTA 199
Db 79 RKMSYIMQDDMLLPHLTVLEAMVSNANLNTENPDVKNLVTETLTALGLMSCSHRTA 138
QY 200 LLSGGQKRLAIAELVNNPPVMEFDEPTSGLDSASCFQVVSMLKSLAQGGRTIICTHQ 259
Db 139 LLSGGQKRLAIAELVNNPPVMEFDEPTSGLDSASCFQVVSMLKSLAQGGRTIICTHQ 198
QY 260 PSAKLFEMFDKLYLSQGCIFKGVVNTNLPYLKGLHCPHTYHNPADFTIEVASGEYGD 319
Db 199 PSAKLFEMFDK-----CIFKGVVNTNLPYLKGLHCPHTYHNPADFTIEVASGEYGD 250
QY 320 LNPMLFRAVONGLCMAEAKSSPEKNEVPAPCPPEVDPIESHTFATSTLTQFCILFK 379
Db 251 LNPMLFRAVONGLCMAEAKSSPEKNEVPAPCPPEVDPIESHTFATSTLTQFCILFK 310
QY 380 RTELSILRDTVLTHLRFMSHVIVGLIGLLYLHIGDDASKVFVNTGCLFFSMLFLMFAAL 439
Db 311 RTELSILRDT-----
QY 440 MPTVLTFPLEMAYFMREHLNLYWYSLKAYLAKTMADVFPQVCPVVCYSIVYVWMTQPAE 499
Db 321 -----VPCPVVCYSIVYVWMTQPAE 340
QY 500 TSLRFLFSALATATATVAOSGLLGAANSLOVATFVGPVTAIPVLLRSGFFVSEKTIPT 559
Db 341 TSLRFLFSALATATATVAOSGLLGAANSLOVATFVGPVTAIPVLLRSGFFVSEKTIPT 400
QY 560 TYLOWSSYLSYVRYGPEGVILTYIGMERGDLTCLEERCPPREPOQSILRALDVEDAKLYMD 619
Db 401 TYLOWSSYLSYVRYGPEGVILTYIGMERGDLTCLEERCPPREPOQSILRALDVEDAKLYMD 460
QY 620 FLVLGIFFLALRLAYLVLYRYR 641
Db 461 FLVLGIFFLALRLAYLVLYRYR 482
RESULT 4
AAM00994
ID AAM00994 standard; Protein: 935 AA.
AC AAM00994;
XX
XX 01-OCT-2001 (first entry)
XX Human bone marrow protein, SEQ ID NO: 495.
XX

PI Patterson C, Gandhi AR, Hafalia AJA, Tribouley CM, Wallia NK;
PI Au-Young J, Walsh RT, Ramkumar J, Lu Y, Lu DAM, Azimzai Y, Lal P;
PI Elliott VS, Nguyen DB, Xu Y, Seilhamer JJ, Borowsky ML, Khan FA;
PI Kearney L, Thangavelu K, Das D, Policky JL;
XX WPI: 2002-205969/26.
DR N-PSDB: AAL44669.
XX
XX New human transporters and ion channel polypeptides for diagnosing,
PT treating or preventing transport, neurological, muscle, immunological
PT and cell proliferative disorders
XX
XX Claim 1; Page 146-147; 230pp; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human transporter and ion channel proteins, designated
CC TRICH-1-TRICH-32. The sequences can be used in the treatment of
CC transport, neurological, muscle, immunological and cell proliferative
CC disorders. The present sequence is a protein of the invention.
XX
XX Sequence 646 AA;
SQ
Query Match 99.88; Score 3339; DB 23; Length 646;
Best Local Similarity 99.78; Pred. No. 3.2e-314;
Matches 644; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MAEKALEAVGCGLPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRSHLPKRSVD 60
Db 1 MAEKALEAVGCGLPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRSHLPKRSVD 60
Qy 61 IEFVELSYVREGPCWKRKGYKTLKCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRE 120
Db 61 IEFVELSYVREGPCWKRKGYKTLKCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRE 120
Qy 121 SGMKGQLLVNRPRELTRFKMSCYIMODDMLLPHLTVLEAMVSNLNTENPDVKNL 180
Db 121 SGMKGQLLVNRPRELTRFKMSCYIMODDMLLPHLTVLEAMVSNLNTENPDVKNL 180
Qy 181 VTEILTALGLMSCSHTTALLSGQKRLAIALELVNPPVPMFDEPTSGLDASCFQV 240
Db 181 VTEILTALGLMSCSHTTALLSGQKRLAIALELVNPPVPMFDEPTSGLDASCFQV 240
Qy 241 SLKSLAQQGRTICTHQSAPKLFEMFDKLYLSQGCIFKGVVTVNLPLKGLGLHCP 300
Db 241 SLKSLAQQGRTICTHQSAPKLFEMFDKLYLSQGCIFKGVVTVNLPLKGLGLHCP 300
Qy 301 TYHPADFIIEVASGEYDGLNPMLEFRVQNGLCMAEKKSSPEKNEVPAPCPPEVD 360
Db 301 TYHPADFIIEVASGEYDGLNPMLEFRVQNGLCMAEKKSSPEKNEVPAPCPPEVD 360
Qy 361 IESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMSHVIVIGLGLYLHIGDDASKV 420
Db 361 IESHTFATSTLTQFCILFKRTFLSILRDTVLTHLRFMSHVIVIGLGLYLHIGDDASKV 420
Qy 421 FNNTCGLFFSNLFLMFAALMPTVLTPLFMAVPMREHLNWTYSLKAYLAKTWADVPFQV 480
Db 421 FNNTCGLFFSNLFLMFAALMPTVLTPLFMAVPMREHLNWTYSLKAYLAKTWADVPFQV 480
Qy 481 VCPVWYCSIVYWTGQPAETSRLLPSALATATATVAQSLGLLIGASNSLQVATFVGPV 540
Db 481 VCPVWYCSIVYWTGQPAETSRLLPSALATATATVAQSLGLLIGASNSLQVATFVGPV 540
Qy 541 TAIPVLLFSGFFVSFKTIPTYLQWSSYLSVYRGFEGVILTIYGMERGDLTCLERCPCR 600
Db 541 TAIPVLLFSGFFVSFKTIPTYLQWSSYLSVYRGFEGVILTIYGMERGDLTCLERCPCR 600
Qy 601 EPOSILRALDVEDAKLYMDFLVIGIFLALRLLAYLVLRVRSKSR 646
Db 601 EPOSILRALDVEDAKLYMDFLVIGIFLALRLLAYLVLRVRSKSR 646

RESULT 2
ABB57112

ID ABB57112 standard; Protein: 666 AA.
XX
XX ABB57112;
XX
XX 07-MAR-2002 (first entry)
XX
XX Mouse ischaemic condition related protein sequence SEQ ID NO:255.
DE
XX
XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW vasospastic ischaemia; ischaemic condition; ischaemic disease.
XX
XX Mus musculus.
OS
XX
XX WO200188188-A2.
PN
XX
XX 22-NOV-2001.
PD
XX
XX 18-MAY-2001; 2001WO-JP04192.
PF
XX
XX 18-MAY-2000; 2000JP-0145977.
PR
XX
XX (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.
PA
XX
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
PI
XX
XX WPI: 2002-034733/04.
DR
XX
XX N-PSDB: ABI99363.
DR
XX
XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT expression levels of particular genes defined in the specification or
PT by determining the expression profile of a gene group comprising these
PT genes -
PT
XX
XX Claim 2; Page 719-722; 2690pp; English.
PS
XX
XX The present invention describes a method for examining ischaemic
CC conditions, comprising measuring the expression levels of particular
CC genes (I) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (I). The method
CC is useful for examining the ischaemic condition (e.g. compressive
CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC expression levels of particular genes (ABI999202 to ABI99912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischaemic condition-improving
CC drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC represent PCR primers for a mouse ischaemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX
XX Sequence 666 AA;
SQ
Query Match 70.38; Score 2353; DB 23; Length 666;
Best Local Similarity 72.28; Pred. No. 1.1e-218;
Matches 451; Conservative 80; Mismatches 90; Indels 4; Gaps 2;
Qy 26 DGAEPVLTTHLKKVENHTEAQRSHLPKRSVDIEFVELSYVREGPCWKRKGYKTL 85
Db 42 DEVETDLLNGLHLKKVNDNNFTAEQRFSLPRAAVNTEFKDLSYVPGCPWKKGYKTL 101
Qy 86 KCLSGKFCRRELIGTMGSPGAGKSTFMNLAGYRESGMKGQILVNGRPRELTRFKMSCY 145
Db 102 KGISGFNSGELVAIMGSPGAGKSTLMNLAGYRETGMKGAVALINGMPDLRCFRKVCY 161
Qy 146 IMODDMLLPHLTVLEAMVSNLNTENPDVKNLTYEILTALGLMSCSHTTALLSGGQ 205
Db 162 IMODDMLLPHLTVOEAMVSAHLKQEGREVMKEILTALGLLPCANTRTRTGSLSGQ 221
Qy 206 RKRLAIALELVNPPVPMFDEPTSGLDASCFQVSLMKSLAQQGRTICTHQSAPKLF 265
Db 222 RKRLAIALELVNPPVPMFDEPTSGLDASCFQVSLMKSLAQQGRTICTHQSAPKLF 281
Qy 266 EMFDKLYLSQGCIFKGVVTVNLPLKGLGLHCPYHNPAIDFIEVASGEYDGLNPMLE 325

Result No.	Score	% Match	Query Length	DB ID	Description
1	3339	99.8	646	23	AA014186 Human transporter
2	2353	70.3	666	23	AB357112 Mouse ischaemic co
3	2299	68.7	935	22	AA178589 Human protein SEQ
4	2299	68.7	935	22	AA000994 Human bone marrow
5	1723	51.5	832	22	AB359544 Drosophila melanog
6	1513	45.2	689	21	AA178980 Silkworm Bm white
7	1322	39.5	689	22	AB361867 Drosophila melanog
8	1241.5	37.1	620	22	AB359648 Drosophila melanog
9	1163.5	34.8	812	22	AB364565 Drosophila melanog
10	1086.5	32.5	609	22	AB364566 Drosophila melanog

ALIGNMENTS

RESULT 1

AAO14186
ID AAO14186 standard; Protein: 646 AA.

AA
AC

AC
XX
AAU4100;

03-MAY-2002 (first entry)

03 JAN 2002 (PRISE ENERGY)
XX
Human transporter and ion channel TRICH-3.
DE

XX Human: transporter and ion channel; TRICH; transport disorder;
KW neurological disorder; muscle disorder; immunological disorder;
KW cell proliferative disorder; neuroprotective; nootropic;
KW cerebroprotective; immunosuppressive; cytostatic; respiratory;
KW gene therapy.

XX
XX
gene therapy.

OS Homo sapiens.

[illegible]

PN WO200204520-A2.

XX

PD 17-JAN-2002.

XX
XX

PF 05-JUL-2001; 2001WO-US21448.

XX

PR 07-JUL-2000; 2000US-216547P.

PR 14-JUL-2000; 2000US-218232P.

PR 21-JUL-2000; 2000US-220112P.
28-JUL-2000; 2000US-231830P.

PR 28-JUL-2000; 2000US-221839P.
yy

XX
DA (INCY -) INCYTE GENOMICS INC

PA (INCY-) INCYTE GENOMICS INC.
XX

Raumann BE Thornton M Di

PT KAUMANN BE, PHILLIPS M, SANTI
PT BURFORD N, GREENE BD, SANTI

PT BUTTOLD N, GREENE BD, SAJJ

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:36:29 ; Search time 14 Seconds
(without alignments)
1913.836 Million cell updates/sec

Title: US-10-072-621-9

Perfect score: 3347

Sequence: 1 MAEKALEAVGCGLGPGAVAM.....FLALRLLAYLVLYRVKRSER 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

rched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3309	98.9	646	ABG4_HUMAN	Q9h172 homo sapien
2	2356	70.4	678	ABG1_HUMAN	P45844 homo sapien
3	2353	70.3	666	ABG1_MOUSE	Q64343 mus musculus
4	879.5	26.3	687	WHIT_DROME	P10090 drosophila
5	806	24.1	679	WHIT_CERCA	Q17320 ceratitidis c
6	794.5	23.7	677	WHIT_LUCCU	Q05360 lucilia cup
7	779	23.3	695	WHIT_ANOGA	Q27256 anopheles g
8	736.5	22.0	709	WHIT_ANOAL	Q16928 anopheles a
9	723.5	21.6	598	YPC3_CAEEL	Q11180 caenorhabdi
10	712	21.3	655	ABG2_HUMAN	Q9unq0 homo sapien
11	667	19.9	666	SCRT_DROME	P45843 drosophila
12	595	17.8	1049	ADP1_YEAST	P23371 saccharomyc
13	585.5	17.5	651	ABG5_HUMAN	Q9h222 homo sapien
14	584	17.4	672	ABG8_RAT	P58428 rattus norv
15	577.5	17.3	652	ABG5_MOUSE	Q99pe8 mus musculu
16	577	17.2	610	YQ5C_CAEEL	Q09466 caenorhabdi
17	565	16.9	1294	YOH5_YEAST	Q08234 saccharomyc
18	563.5	16.8	652	ABG5_RAT	Q99pe7 rattus norv
19	563.5	16.8	673	ABG8_HUMAN	Q9h221 homo sapien
20	557.5	16.7	673	ABG8_MOUSE	Q9dbm0 mus musculu
21	531.5	15.9	650	ABG3_MOUSE	Q99pe81 mus musculu
22	483	14.4	1333	YN99_YEAST	P53756 saccharomyc
23	472	14.1	1529	PDRE_YEAST	Q04182 saccharomyc
24	469	14.0	1530	BFP1_SCHPO	P41820 schizosacch
25	466.5	13.9	1511	PDRC_YEAST	Q02785 saccharomyc
26	462.5	13.8	1564	PDRA_YEAST	P51533 saccharomyc
27	459	13.7	1501	SNQ2_YEAST	P32568 saccharomyc
28	455	13.6	675	BROW_DROME	P12428 drosophila
29	445.5	13.3	1511	PDRE_YEAST	P33302 saccharomyc
30	437.5	13.1	1501	CDR3_CANAL	Q04676 candida alb
31	429	12.8	1490	CDR4_CANAL	P42690 candida alb
32	424.5	12.7	1499	CDR2_CANAL	P78595 candida alb
33	413.5	12.4	1501	CDR1_CANAL	P43071 candida alb

RESULT 1

ID	ABG4_HUMAN	STANDARD	PRT	646 AA
AC	Q9HJ72			
DT	15-JUN-2002	(Rel. 41, Created)		
DT	15-JUN-2002	(Rel. 41, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	ATP-binding cassette, sub-family G, member 4.			
GN	ABCG4 OR WHITE2			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid:9606			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21518231; PubMed=11606068;			
RA	Engel T., Lorkowski S., Lueken A., Rust S., Schlueter B., Berger G.,			
RA	Cullen P., Assmann G.;			
RT	"The human ABCG4 gene is regulated by oxysterols and retinoids in			
RT	monocyte-derived macrophages."			
RL	Biochem. Biophys. Res. Commun. 288:483-488(2001).			
RN	[2]			
RP	SEQUENCE OF 20-646 FROM N.A.			
RC	TISSUE=Dorsal root ganglion;			
RA	Oldfield S., Lowry C.A., Lightman S.L.;			
RT	"Cloning and expression of a novel mammalian white family			
RT	ABC transporter: WHITE2."			
RL	Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.			
CC	!- FUNCTION: May be involved in macrophage lipid homeostasis.			
CC	!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).			
CC	!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; AJ300465; CAC17140.1;			
DR	Genew; HGNC:13884; ABCG4.			
DR	InterPro; IPR003593; AAA_ATPase.			
DR	InterPro; IPR003439; ABC_transportr.			
DR	Pfam; PF00005; ABC_tran.2			
DR	ProDom; PD0000006; ABC_transportr; 1.			
DR	SMART; SM00382; AAA; 1.			
DR	PROSITE; PS00211; ABC_TRANSPORTER; 1.			
KW	ATP-binding; Glycoprotein; Transmembrane; Transport.			
FT	DOMAIN 1 393			
FT	TRANSMEM 394 414			
FT	DOMAIN 415 425			
FT	TRANSMEM 426 446			
FT	DOMAIN 447 472			
FT	TRANSMEM 473 493			
FT	DOMAIN 1 393			
FT	TRANSMEM 394 414			
FT	DOMAIN 415 425			
FT	TRANSMEM 426 446			
FT	DOMAIN 447 472			
FT	TRANSMEM 473 493			

Q24739 drosophila
P40550 saccharomyc
P50332 rhizobium g
Q58903 methanococ
P08720 rhizobium l
P54537 bacillus su
Q05067 rhizobium s
Q05067 rhizobium s
P10346 escherichia
P75612 mycoplasma
Q9wy17 thermotoga
P44986 haemophilus

ALIGNMENTS

```

FT DOMAIN 494 503 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 504 524 4 (POTENTIAL).
FT DOMAIN 525 532 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 533 553 5 (POTENTIAL).
FT DOMAIN 554 617 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 618 638 6 (POTENTIAL).
FT DOMAIN 639 646 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 102 109 ATP (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 646 AA: 71895 MW: 9CCECE6E150772611 CRC64;

Query Match 98.9% Score 3309; DB 1; Length 646;
Best Local Similarity 98.9%; Pred. No. 3; le-214;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MAEKALEAVGCGLGPGAVAMVTELDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAVD 60
Db 1 MAEKALEAVGCGLGPGAVAMVTELDGAEPPVLTTHLKKVENHTEAQRSHLPKRSAVD 60
QY 61 IEFVELSYVREGPCWRKRGYKTLKLSGKFCRRELIGIMGPGSGAGKSTFNMILAGYRE 120
61 IEFVELSYVREGPCWRKRGYKTLKLSGKFCRRELIGIMGPGSGAGKSTFNMILAGYRE 120
121 SGMKQOILVNGPRELFRKMSYIMQDDMLLPHLTVLEAMWVSANLNTENPDVKNDL 180
121 SGMKQOILVNGPRELFRKMSYIMQDDMLLPHLTVLEAMWVSANLNTENPDVKNDL 180
QY 181 VTEILTALGMSCHTRTALLSGGQRKLATALELVNPNPVFFDEPTSGLDASCFQV 240
181 VTEILTALGMSCHTRTALLSGGQRKLATALELVNPNPVFFDEPTSGLDASCFQV 240
QY 241 SLMSLAQGGRTIICTIHOPSAKLFEMEDKLYLSQGGCIFKGVVNTLIPYLKGLGLHCP 300
241 SLMSLAQGGRTIICTIHOPSAKLFEMEDKLYLSQGGCIFKGVVNTLIPYLKGLGLHCP 300
QY 301 TYHNPADRIIEVASGEYGLDPLFRAVONGLCMAEKKSSPEKNEVPAPCPPEVD 360
301 TYHNPADRIIEVASGEYGLDPLFRAVONGLCMAEKKSSPEKNEVPAPCPPEVD 360
QY 361 IESHFTFATSTLQFLEKRTFLSLRDTVLTHLFMSHVIVGLVGLLHIGDASKV 420
361 IESHFTFATSTLQFLEKRTFLSLRDTVLTHLFMSHVIVGLVGLLHIGDASKV 420
QY 421 FNNTCGLFFSMLFLMFAALMPTVLTFFLEMAVFMREHLNYWYSLKAYLAKTMADVPFQV 480
421 FNNTCGLFFSMLFLMFAALMPTVLTFFLEMAVFMREHLNYWYSLKAYLAKTMADVPFQV 480
QY 481 VCPVVCYSIVVMTGQPAETSRFLFSALATATATVAOSGLGLLGAASNSQVATFVGPV 540
481 VCPVVCYSIVVMTGQPAETSRFLFSALATATATVAOSGLGLLGAASNSQVATFVGPV 540
QY 541 TAIPVLLFSGFFSVFSTKPTIYQWSSYLSYVRYGEGVILTYGMERGLDTCLEERCPCR 600
541 TAIPVLLFSGFFSVFSTKPTIYQWSSYLSYVRYGEGVILTYGMERGLDTCLEERCPCR 600
QY 601 EPQSTILRALDVEDAKLYMDFLVGLVGIFFLALRLAYLVLYRYVKSR 646
601 EPQSTILRALDVEDAKLYMDFLVGLVGIFFLALRLAYLVLYRYVKSR 646

```

RESULT 2

```

ABG1_HUMAN STANDARD; PRT; 678 AA.
ID ABG1_HUMAN Q9BXL0; Q9BXL1; Q9BXL2; Q9BXL3; Q9BXL4; Q9BXL6; Q9BXL7;
AC P45844; Q9BXL0; Q9BXL1; Q9BXL2; Q9BXL3; Q9BXL4; Q9BXL6; Q9BXL7;
AC Q9BXL8; Q9BXL9;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (White protein homolog)
DE (ATP-binding cassette transporter 8).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
RC TISSUE=Retina;
RX MEDLINE=96256850; PubMed=8659545;
RA Chen H.M., Rossier C., Lalliot M.D., Lynn A., Chakravarti A.,
Perrin G., Antonarakis S.E.;
RT "Cloning of the cDNA for a human homologue of the Drosophila white
gene and mapping to chromosome 21q22.3.";
RL Am. J. Hum. Genet. 59:66-75(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
Ramsey J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
Lehrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20408883; PubMed=10950923;
RA Berry A., Scott H.S., Kudoh J., Tallor I., Korostishevsky M.,
Wattenhofer M., Guipponi M., Bartas C., Rossier C., Shibuya K.,
Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,
Antonarakis S.E., Bonne-Tamir B.;
RT "Refined localization of autosomal recessive nonsyndromic deafness
DFNB10 locus using 34 novel microsatellite markers, genomic
structures, and exclusion of six known genes in the region.";
RL Genomics 68:22-29(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=21192304; PubMed=11279031;
RA Porsch-Oezueruemaz M., Langmann T., Heimerl S., Borsukova H.,
Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;
RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor
of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
expression and a modulator of cellular lipid efflux.";
RL J. Biol. Chem. 276:12427-12433(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).
RX MEDLINE=21092576; PubMed=11162488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
Assmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [6]
RP SEQUENCE OF 33-678 FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97186700; PubMed=9034316;
RA Croop J.M., Filler G.E., Fletcher J.A., Lux M.L., Raab E.,
Goldenson D., Arciniegas S., Son D., Wu R.;
RT "Isolation and characterization of a mammalian homologue of the
Drosophila white gene.";
RL Gene 185:77-85(1997).
RN [7]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE=20261604; PubMed=10799558;
RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
lipid-loaded macrophages. A transcriptional role for specific
oxysterols.";
RL J. Biol. Chem. 275:14700-14707(2000).

```

RN [8]
 RP INDUCTION, AND PROBABLE FUNCTION.
 RX MEDLINE=2010556; PubMed=10639163;
 RA Klucken J., Buechler C., Orso E., Kaminski W.E.,
 RA Porsch-Oerzueruem M., Liebisch G., Kapinsky M., Diederich W.,
 RA Drobniak W., Dean M., Alilkin R., Schmitz G.;
 RT "ABCG1 (ABC8), the human homolog of the phospholipid white gene, is a
 RT regulator of macrophage cholesterol and phospholipid transport.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).
 RN [9]
 RP REVIEW
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is
 CC an active component of the macrophage lipid export complex. Could
 CC also be involved in intracellular lipid transport processes. The
 CC role in cellular lipid homeostasis may not be limited to
 CC macrophages.
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners
 CC of the ABCG subfamily.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC Predominantly localized in the intracellular compartments mainly
 CC associated with the endoplasmic reticulum (ER) and Golgi
 CC membranes.
 CC -!- ALTERNATIVE PRODUCTS: At least 7 isoforms: 1 (shown here), 2/J,
 CC 3/ABDE, 4/G, 5/F, 6/HI and 7/C; are produced by alternative
 CC splicing and alternative initiation.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.
 CC -!- INDUCTION: Strongly induced in monocyte-derived macrophages during
 CC cholesterol influx. Conversely, mRNA and protein expression are
 CC suppressed by lipid efflux. Induction is mediated by the
 CC liver X receptor/retinoid X receptor (LXR/RXR) pathway.
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X91249; CAA62631.1; ALT_INIT.
 DR EMBL; AP001746; BAA95530.1; ALT_INIT.
 DR EMBL; AB038161; BAB13728.2; ALT_INIT.
 DR EMBL; AJ289137; CAC00730.1; JOINED.
 DR EMBL; AJ289138; CAC00730.1; JOINED.
 DR EMBL; AJ289139; CAC00730.1; JOINED.
 DR EMBL; AJ289140; CAC00730.1; JOINED.
 DR EMBL; AJ289141; CAC00730.1; JOINED.
 DR EMBL; AJ289142; CAC00730.1; JOINED.
 DR EMBL; AJ289143; CAC00730.1; JOINED.
 DR EMBL; AJ289144; CAC00730.1; JOINED.
 DR EMBL; AJ289145; CAC00730.1; JOINED.
 DR EMBL; AJ289146; CAC00730.1; JOINED.
 DR EMBL; AJ289147; CAC00730.1; JOINED.
 DR EMBL; AJ289148; CAC00730.1; JOINED.
 DR EMBL; AJ289149; CAC00730.1; JOINED.
 DR EMBL; AJ289150; CAC00730.1; JOINED.
 DR EMBL; AJ289151; CAC00730.1; JOINED.
 DR EMBL; AF323658; AAK28836.1; JOINED.
 DR EMBL; AF323664; AAK28836.1; JOINED.
 DR EMBL; AF323665; AAK28836.1; JOINED.
 DR EMBL; AF323667; AAK28836.1; JOINED.
 DR EMBL; AF323668; AAK28836.1; JOINED.
 DR EMBL; AF323669; AAK28836.1; JOINED.
 DR EMBL; AF323670; AAK28836.1; JOINED.
 DR EMBL; AF323671; AAK28836.1; JOINED.
 DR EMBL; AF323672; AAK28836.1; JOINED.
 DR EMBL; AF323673; AAK28836.1; JOINED.
 DR EMBL; AF323674; AAK28836.1; JOINED.
 DR EMBL; AF323675; AAK28836.1; JOINED.
 DR EMBL; AF323676; AAK28836.1; JOINED.
 DR EMBL; AF323677; AAK28836.1; JOINED.
 DR EMBL; AF323678; AAK28836.1; JOINED.
 DR EMBL; AF323679; AAK28836.1; JOINED.
 DR EMBL; AF323680; AAK28836.1; JOINED.
 DR EMBL; AF323681; AAK28836.1; JOINED.
 DR EMBL; AF323682; AAK28836.1; JOINED.
 DR EMBL; AF323683; AAK28836.1; JOINED.
 DR EMBL; AF323684; AAK28836.1; JOINED.
 DR EMBL; AF323685; AAK28836.1; JOINED.
 DR EMBL; AF323686; AAK28836.1; JOINED.
 DR EMBL; AF323687; AAK28836.1; JOINED.
 DR EMBL; AF323688; AAK28836.1; JOINED.
 DR EMBL; AF323689; AAK28836.1; JOINED.
 DR EMBL; AF323690; AAK28836.1; JOINED.
 DR EMBL; AF323691; AAK28836.1; JOINED.
 DR EMBL; AF323692; AAK28836.1; JOINED.
 DR EMBL; AF323693; AAK28836.1; JOINED.
 DR EMBL; AF323694; AAK28836.1; JOINED.
 DR EMBL; AF323695; AAK28836.1; JOINED.
 DR EMBL; AF323696; AAK28836.1; JOINED.
 DR EMBL; AF323697; AAK28836.1; JOINED.
 DR EMBL; AF323698; AAK28836.1; JOINED.
 DR EMBL; AF323699; AAK28836.1; JOINED.
 DR EMBL; AF323700; AAK28836.1; JOINED.
 DR EMBL; AF323701; AAK28836.1; JOINED.
 DR EMBL; AF323702; AAK28836.1; JOINED.
 DR EMBL; AF323703; AAK28836.1; JOINED.
 DR EMBL; AF323704; AAK28836.1; JOINED.
 DR EMBL; AF323705; AAK28836.1; JOINED.
 DR EMBL; AF323706; AAK28836.1; JOINED.
 DR EMBL; AF323707; AAK28836.1; JOINED.
 DR EMBL; AF323708; AAK28836.1; JOINED.
 DR EMBL; AF323709; AAK28836.1; JOINED.
 DR EMBL; AF323710; AAK28836.1; JOINED.
 DR EMBL; AF323711; AAK28836.1; JOINED.
 DR EMBL; AF323712; AAK28836.1; JOINED.
 DR EMBL; AF323713; AAK28836.1; JOINED.
 DR EMBL; AF323714; AAK28836.1; JOINED.
 DR EMBL; AF323715; AAK28836.1; JOINED.
 DR EMBL; AF323716; AAK28836.1; JOINED.
 DR EMBL; AF323717; AAK28836.1; JOINED.
 DR EMBL; AF323718; AAK28836.1; JOINED.
 DR EMBL; AF323719; AAK28836.1; JOINED.
 DR EMBL; AF323720; AAK28836.1; JOINED.
 DR EMBL; AF323721; AAK28836.1; JOINED.
 DR EMBL; AF323722; AAK28836.1; JOINED.
 DR EMBL; AF323723; AAK28836.1; JOINED.
 DR EMBL; AF323724; AAK28836.1; JOINED.
 DR EMBL; AF323725; AAK28836.1; JOINED.
 DR EMBL; AF323726; AAK28836.1; JOINED.
 DR EMBL; AF323727; AAK28836.1; JOINED.
 DR EMBL; AF323728; AAK28836.1; JOINED.
 DR EMBL; AF323729; AAK28836.1; JOINED.
 DR EMBL; AF323730; AAK28836.1; JOINED.
 DR EMBL; AF323731; AAK28836.1; JOINED.
 DR EMBL; AF323732; AAK28836.1; JOINED.
 DR EMBL; AF323733; AAK28836.1; JOINED.
 DR EMBL; AF323734; AAK28836.1; JOINED.
 DR EMBL; AF323735; AAK28836.1; JOINED.
 DR EMBL; AF323736; AAK28836.1; JOINED.
 DR EMBL; AF323737; AAK28836.1; JOINED.
 DR EMBL; AF323738; AAK28836.1; JOINED.
 DR EMBL; AF323739; AAK28836.1; JOINED.
 DR EMBL; AF323740; AAK28836.1; JOINED.
 DR EMBL; AF323741; AAK28836.1; JOINED.
 DR EMBL; AF323742; AAK28836.1; JOINED.
 DR EMBL; AF323743; AAK28836.1; JOINED.
 DR EMBL; AF323744; AAK28836.1; JOINED.
 DR EMBL; AF323745; AAK28836.1; JOINED.
 DR EMBL; AF323746; AAK28836.1; JOINED.
 DR EMBL; AF323747; AAK28836.1; JOINED.
 DR EMBL; AF323748; AAK28836.1; JOINED.
 DR EMBL; AF323749; AAK28836.1; JOINED.
 DR EMBL; AF323750; AAK28836.1; JOINED.
 DR EMBL; AF323751; AAK28836.1; JOINED.
 DR EMBL; AF323752; AAK28836.1; JOINED.
 DR EMBL; AF323753; AAK28836.1; JOINED.
 DR EMBL; AF323754; AAK28836.1; JOINED.
 DR EMBL; AF323755; AAK28836.1; JOINED.
 DR EMBL; AF323756; AAK28836.1; JOINED.
 DR EMBL; AF323757; AAK28836.1; JOINED.
 DR EMBL; AF323758; AAK28836.1; JOINED.
 DR EMBL; AF323759; AAK28836.1; JOINED.
 DR EMBL; AF323760; AAK28836.1; JOINED.
 DR EMBL; AF323761; AAK28836.1; JOINED.
 DR EMBL; AF323762; AAK28836.1; JOINED.
 DR EMBL; AF323763; AAK28836.1; JOINED.
 DR EMBL; AF323764; AAK28836.1; JOINED.
 DR EMBL; AF323765; AAK28836.1; JOINED.
 DR EMBL; AF323766; AAK28836.1; JOINED.
 DR EMBL; AF323767; AAK28836.1; JOINED.
 DR EMBL; AF323768; AAK28836.1; JOINED.
 DR EMBL; AF323769; AAK28836.1; JOINED.
 DR EMBL; AF323770; AAK28836.1; JOINED.
 DR EMBL; AF323771; AAK28836.1; JOINED.
 DR EMBL; AF323772; AAK28836.1; JOINED.
 DR EMBL; AF323773; AAK28836.1; JOINED.
 DR EMBL; AF323774; AAK28836.1; JOINED.
 DR EMBL; AF323775; AAK28836.1; JOINED.
 DR EMBL; AF323776; AAK28836.1; JOINED.
 DR EMBL; AF323777; AAK28836.1; JOINED.
 DR EMBL; AF323778; AAK28836.1; JOINED.
 DR EMBL; AF323779; AAK28836.1; JOINED.
 DR EMBL; AF323780; AAK28836.1; JOINED.
 DR EMBL; AF323781; AAK28836.1; JOINED.
 DR EMBL; AF323782; AAK28836.1; JOINED.
 DR EMBL; AF323783; AAK28836.1; JOINED.
 DR EMBL; AF323784; AAK28836.1; JOINED.
 DR EMBL; AF323785; AAK28836.1; JOINED.
 DR EMBL; AF323786; AAK28836.1; JOINED.
 DR EMBL; AF323787; AAK28836.1; JOINED.
 DR EMBL; AF323788; AAK28836.1; JOINED.
 DR EMBL; AF323789; AAK28836.1; JOINED.
 DR EMBL; AF323790; AAK28836.1; JOINED.
 DR EMBL; AF323791; AAK28836.1; JOINED.
 DR EMBL; AF323792; AAK28836.1; JOINED.
 DR EMBL; AF323793; AAK28836.1; JOINED.
 DR EMBL; AF323794; AAK28836.1; JOINED.
 DR EMBL; AF323795; AAK28836.1; JOINED.
 DR EMBL; AF323796; AAK28836.1; JOINED.
 DR EMBL; AF323797; AAK28836.1; JOINED.
 DR EMBL; AF323798; AAK28836.1; JOINED.
 DR EMBL; AF323799; AAK28836.1; JOINED.
 DR EMBL; AF323800; AAK28836.1; JOINED.
 DR EMBL; AF323801; AAK28836.1; JOINED.
 DR EMBL; AF323802; AAK28836.1; JOINED.
 DR EMBL; AF323803; AAK28836.1; JOINED.
 DR EMBL; AF323804; AAK28836.1; JOINED.
 DR EMBL; AF323805; AAK28836.1; JOINED.
 DR EMBL; AF323806; AAK28836.1; JOINED.
 DR EMBL; AF323807; AAK28836.1; JOINED.
 DR EMBL; AF323808; AAK28836.1; JOINED.
 DR EMBL; AF323809; AAK28836.1; JOINED.
 DR EMBL; AF323810; AAK28836.1; JOINED.
 DR EMBL; AF323811; AAK28836.1; JOINED.
 DR EMBL; AF323812; AAK28836.1; JOINED.
 DR EMBL; AF323813; AAK28836.1; JOINED.
 DR EMBL; AF323814; AAK28836.1; JOINED.
 DR EMBL; AF323815; AAK28836.1; JOINED.
 DR EMBL; AF323816; AAK28836.1; JOINED.
 DR EMBL; AF323817; AAK28836.1; JOINED.
 DR EMBL; AF323818; AAK28836.1; JOINED.
 DR EMBL; AF323819; AAK28836.1; JOINED.
 DR EMBL; AF323820; AAK28836.1; JOINED.
 DR EMBL; AF323821; AAK28836.1; JOINED.
 DR EMBL; AF323822; AAK28836.1; JOINED.
 DR EMBL; AF323823; AAK28836.1; JOINED.
 DR EMBL; AF323824; AAK28836.1; JOINED.
 DR EMBL; AF323825; AAK28836.1; JOINED.
 DR EMBL; AF323826; AAK28836.1; JOINED.
 DR EMBL; AF323827; AAK28836.1; JOINED.
 DR EMBL; AF323828; AAK28836.1; JOINED.
 DR EMBL; AF323829; AAK28836.1; JOINED.
 DR EMBL; AF323830; AAK28836.1; JOINED.
 DR EMBL; AF323831; AAK28836.1; JOINED.
 DR EMBL; AF323832; AAK28836.1; JOINED.
 DR EMBL; AF323833; AAK28836.1; JOINED.
 DR EMBL; AF323834; AAK28836.1; JOINED.
 DR EMBL; AF323835; AAK28836.1; JOINED.
 DR EMBL; AF323836; AAK28836.1; JOINED.
 DR EMBL; AF323837; AAK28836.1; JOINED.
 DR EMBL; AF323838; AAK28836.1; JOINED.
 DR EMBL; AF323839; AAK28836.1; JOINED.
 DR EMBL; AF323840; AAK28836.1; JOINED.
 DR EMBL; AF323841; AAK28836.1; JOINED.
 DR EMBL; AF323842; AAK28836.1; JOINED.
 DR EMBL; AF323843; AAK28836.1; JOINED.
 DR EMBL; AF323844; AAK28836.1; JOINED.
 DR EMBL; AF323845; AAK28836.1; JOINED.
 DR EMBL; AF323846; AAK28836.1; JOINED.
 DR EMBL; AF323847; AAK28836.1; JOINED.
 DR EMBL; AF323848; AAK28836.1; JOINED.
 DR EMBL; AF323849; AAK28836.1; JOINED.
 DR EMBL; AF323850; AAK28836.1; JOINED.
 DR EMBL; AF323851; AAK28836.1; JOINED.
 DR EMBL; AF323852; AAK28836.1; JOINED.
 DR EMBL; AF323853; AAK28836.1; JOINED.
 DR EMBL; AF323854; AAK28836.1; JOINED.
 DR EMBL; AF323855; AAK28836.1; JOINED.
 DR EMBL; AF323856; AAK28836.1; JOINED.
 DR EMBL; AF323857; AAK28836.1; JOINED.
 DR EMBL; AF323858; AAK28836.1; JOINED.
 DR EMBL; AF323859; AAK28836.1; JOINED.
 DR EMBL; AF323860; AAK28836.1; JOINED.
 DR EMBL; AF323861; AAK28836.1; JOINED.
 DR EMBL; AF323862; AAK28836.1; JOINED.
 DR EMBL; AF323863; AAK28836.1; JOINED.
 DR EMBL; AF323864; AAK28836.1; JOINED.
 DR EMBL; AF323865; AAK28836.1; JOINED.
 DR EMBL; AF323866; AAK28836.1; JOINED.
 DR EMBL; AF323867; AAK28836.1; JOINED.
 DR EMBL; AF323868; AAK28836.1; JOINED.
 DR EMBL; AF323869; AAK28836.1; JOINED.
 DR EMBL; AF323870; AAK28836.1; JOINED.
 DR EMBL; AF323871; AAK28836.1; JOINED.
 DR EMBL; AF323872; AAK28836.1; JOINED.
 DR EMBL; AF323873; AAK28836.1; JOINED.
 DR EMBL; AF323874; AAK28836.1; JOINED.
 DR EMBL; AF323875; AAK28836.1; JOINED.
 DR EMBL; AF323876; AAK28836.1; JOINED.
 DR EMBL; AF323877; AAK28836.1; JOINED.
 DR EMBL; AF323878; AAK28836.1; JOINED.
 DR EMBL; AF323879; AAK28836.1; JOINED.
 DR EMBL; AF323880; AAK28836.1; JOINED.
 DR EMBL; AF323881; AAK28836.1; JOINED.
 DR EMBL; AF323882; AAK28836.1; JOINED.
 DR EMBL; AF323883; AAK28836.1; JOINED.
 DR EMBL; AF323884; AAK28836.1; JOINED.
 DR EMBL; AF323885; AAK28836.1; JOINED.
 DR EMBL; AF323886; AAK28836.1; JOINED.
 DR EMBL; AF323887; AAK28836.1; JOINED.
 DR EMBL; AF323888; AAK28836.1; JOINED.
 DR EMBL; AF323889; AAK28836.1; JOINED.
 DR EMBL; AF323890; AAK28836.1; JOINED.
 DR EMBL; AF323891; AAK28836.1; JOINED.
 DR EMBL; AF323892; AAK28836.1; JOINED.
 DR EMBL; AF323893; AAK28836.1; JOINED.
 DR EMBL; AF323894; AAK28836.1; JOINED.
 DR EMBL; AF323895; AAK28836.1; JOINED.
 DR EMBL; AF323896; AAK28836.1; JOINED.
 DR EMBL; AF323897; AAK28836.1; JOINED.
 DR EMBL; AF323898; AAK28836.1; JOINED.
 DR EMBL; AF323899; AAK28836.1; JOINED.
 DR EMBL; AF323900; AAK28836.1; JOINED.
 DR EMBL; AF323901; AAK28836.1; JOINED.
 DR EMBL; AF323902; AAK28836.1; JOINED.
 DR EMBL; AF323903; AAK28836.1; JOINED.
 DR EMBL; AF323904; AAK28836.1; JOINED.
 DR EMBL; AF323905; AAK28836.1; JOINED.
 DR EMBL; AF323906; AAK28836.1; JOINED.
 DR EMBL; AF323907; AAK28836.1; JOINED.
 DR EMBL; AF323908; AAK28836.1; JOINED.
 DR EMBL; AF323909; AAK28836.1; JOINED.
 DR EMBL; AF323910; AAK28836.1; JOINED.
 DR EMBL; AF323911; AAK28836.1; JOINED.
 DR EMBL; AF323912; AAK28836.1; JOINED.
 DR EMBL; AF323913; AAK28836.1; JOINED.
 DR EMBL; AF323914; AAK28836.1; JOINED.
 DR EMBL; AF323915; AAK28836.1; JOINED.
 DR EMBL; AF323916; AAK28836.1; JOINED.
 DR EMBL; AF323917; AAK28836.1; JOINED.
 DR EMBL; AF323918; AAK28836.1; JOINED.
 DR EMBL; AF323919; AAK28836.1; JOINED.
 DR EMBL; AF323920; AAK28836.1; JOINED.
 DR EMBL; AF323921; AAK28836.1; JOINED.
 DR EMBL; AF323922; AAK28836.1; JOINED.
 DR EMBL; AF323923; AAK28836.1; JOINED.
 DR EMBL; AF323924; AAK28836.1; JOINED.
 DR EMBL; AF323925; AAK28836.1; JOINED.
 DR EMBL; AF323926; AAK28836.1; JOINED.
 DR EMBL; AF323927; AAK28836.1; JOINED.
 DR EMBL; AF323928; AAK28836.1; JOINED.
 DR EMBL; AF323929; AAK28836.1; JOINED.
 DR EMBL; AF323930; AAK28836.1; JOINED.
 DR EMBL; AF323931; AAK28836.1; JOINED.
 DR EMBL; AF323932; AAK28836.1; JOINED.
 DR EMBL; AF323933; AAK28836.1; JOINED.
 DR EMBL; AF323934; AAK28836.1; JOINED.
 DR EMBL; AF323935; AAK28836.1; JOINED.
 DR EMBL; AF323936; AAK28836.1; JOINED.
 DR EMBL; AF323937; AAK28836.1; JOINED.
 DR EMBL; AF323938; AAK28836.1; JOINED.
 DR EMBL; AF323939; AAK28836.1; JOINED.
 DR EMBL; AF323940; AAK28836.1; JOINED.
 DR EMBL; AF323941; AAK28836.1; JOINED.
 DR EMBL; AF323942; AAK28836.1; JOINED.
 DR EMBL; AF323943; AAK28836.1; JOINED.
 DR EMBL; AF323944; AAK28836.1; JOINED.
 DR EMBL; AF323945; AAK28836.1; JOINED.
 DR EMBL; AF323946; AAK28836.1; JOINED.
 DR EMBL; AF323947; AAK28836.1; JOINED.
 DR EMBL; AF323948; AAK28836.1; JOINED.
 DR EMBL; AF323949; AAK28836.1; JOINED.
 DR EMBL; AF323950; AAK28836.1; JOINED.
 DR EMBL; AF323951; AAK28836.1; JOINED.
 DR EMBL; AF323952; AAK28836.1; JOINED.
 DR EMBL; AF323953; AAK28836.1; JOINED.
 DR EMBL; AF323954; AAK28836.1; JOINED.
 DR EMBL; AF323955; AAK28836.1; JOINED.
 DR EMBL; AF323956; AAK28836.1; JOINED.
 DR EMBL; AF323957; AAK28836.1; JOINED.
 DR EMBL; AF323958; AAK28836.1; JOINED.
 DR EMBL; AF323959; AAK28836.1; JOINED.
 DR EMBL; AF323960; AAK28836.1; JOINED.
 DR EMBL; AF323961; AAK28836.1; JOINED.
 DR EMBL; AF323962; AAK28836.1; JOINED.
 DR EMBL; AF323963; AAK28836.1; JOINED.
 DR EMBL; AF323964; AAK28836.1; JOINED.
 DR EMBL; AF323965; AAK28836.1; JOINED.
 DR EMBL; AF323966; AAK28836.1; JOINED.
 DR EMBL; AF323967; AAK28836.1; JOINED.
 DR EMBL; AF323968; AAK28836.1; JOINED.
 DR EMBL; AF323969; AAK28836.1; JOINED.
 DR EMBL; AF323970; AAK28836.1; JOINED.
 DR EMBL; AF323971; AAK28836.1; JOINED.
 DR EMBL; AF323972; AAK28836.1; JOINED.
 DR EMBL; AF323973; AAK28836.1; JOINED.
 DR EMBL; AF323974; AAK28836.1; JOINED.
 DR EMBL; AF323975; AAK28836.1; JOINED.
 DR EMBL; AF323976; AAK28836.1; JOINED.
 DR EMBL; AF323977; AAK28836.1; JOINED.
 DR EMBL; AF323978; AAK28836.1; JOINED.
 DR EMBL; AF323979; AAK28836.1; JOINED.
 DR EMBL; AF323980; AAK28836.1; JOINED.
 DR EMBL; AF323981; AAK28836.1; JOINED.
 DR EMBL; AF323982; AAK28836.1; JOINED.
 DR EMBL; AF323983; AAK28836.1; JOINED.
 DR EMBL; AF323984; AAK28836.1; JOINED.
 DR EMBL; AF323985; AAK28836.1; JOINED.
 DR EMBL; AF323986; AAK28836.1; JOINED.
 DR EMBL; AF323987; AAK28836.1; JOINED.
 DR EMBL; AF323988; AAK28836.1; JOINED.
 DR EMBL; AF323989; AAK28836.1; JOINED.
 DR EMBL; AF323990; AAK28836.1; JOINED.
 DR EMBL; AF323991; AAK28836.1; JOINED.
 DR EMBL; AF323992; AAK28836.1; JOINED.
 DR EMBL; AF323993; AAK28836.1; JOINED.
 DR EMBL; AF323994; AAK28836.1; JOINED.
 DR EMBL; AF323995; AAK28836.1; JOINED.
 DR EMBL; AF323996; AAK28836.1; JOINED.
 DR EMBL; AF323997; AAK28836.1; JOINED.
 DR EMBL; AF323998; AAK28836.1; JOINED.
 DR EMBL; AF323999; AAK28836.1; JOINED.
 DR EMBL; AF324000; AAK28836.1; JOINED.
 DR EMBL; AF324001; AAK28836.1; JOINED.
 DR EMBL; AF324002; AAK28836.1; JOINED.
 DR EMBL; AF324003; AAK28836.1; JOINED.
 DR EMBL; AF324004; AAK28836.1; JOINED.
 DR EMBL; AF324005; AAK28836.1; JOINED.
 DR EMBL; AF324006; AAK28836.1; JOINED.
 DR EMBL; AF324007; AAK28836.1; JOINED.
 DR EMBL; AF324008; AAK28836.1; JOINED.
 DR EMBL; AF324009; AAK28836.1; JOINED.
 DR EMBL; AF324010; AAK28836.1; JOINED.
 DR EMBL; AF324011; AAK28836.1; JOINED.
 DR EMBL; AF324012; AAK28836.1; JOINED.
 DR EMBL; AF324013; AAK28836.1; JOINED.
 DR EMBL; AF324014; AAK28836.1; JOINED.
 DR EMBL; AF324015; AAK28836.1; JOINED.
 DR EMBL; AF324016; AAK28836.1; JOINED.
 DR EMBL; AF324017; AAK28836.1; JOINED.
 DR EMBL; AF324018; AAK28836.1; JOINED.
 DR EMBL; AF324019; AAK28836.1; JOINED.
 DR EMBL; AF324020; AAK28836.1; JOINED.
 DR EMBL; AF324021; AAK28836.1; JOINED.
 DR EMBL; AF324022; AAK28836.1; JOINED.
 DR EMBL; AF324023; AAK28836.1; JOIN

```

Db 391 MEGCHSFSASGLTOPCILFKRTFLSMDSDSVLTHLRITSHIGLIGLLGLLYGIGNEAK 450
Qy 420 VFNNTGCLFFESMLFMAALMPTVLTPLFMAVFNREHLNWSLKAYLAKTMADVPFQ 479
Db 451 VLSNCGFFESMLFMAALMPTVLTPLFMAVFNREHLNWSLKAYLAKTMADVPFQ 510
Qy 480 VCPVYCSIVYVMTGQPAETSRLFLFSALATATVAQSLGLITGAASLSQVATFVGP 539
Db 511 IMFVAYCSIVYVMTSQSDAVRFVLFALGTWMSLVAQSLGLITGAASLSQVATFVGP 570
Qy 540 VTAIPVLLFSGFFSVFRTPIPYLQMSVSVYRGFEGVILTIYGMERGDITC-LEERCP 598
Db 571 VTAIPVLLFSGFFSVFRTPIPYLQMSVSVYRGFEGVILTIYGMERGDITC-LEERCP 630
Qy 599 FREPQSIILRALDVEDAKLYMDFVLGIFFLALRLAYLVLRYVKSR 646
Db 631 FOKSAILRELDVENAKLYLDFIVLGIFFLALRLAYLVLRYVKSR 678

RESULT 3
1. MOUSE
ABGI_MOUSE STANDARD: PRT: 666 AA.
Q64343;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 1 (white protein homolog)
DE (ATP-binding cassette transporter 8).
GN ABCG1 OR ABC8 OR WH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97186700; PubMed=9034316;
RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,
RA Goldenson D., Son D., Arciniegas S., Wu R.;
RT "Isolation and characterization of a mammalian homolog of the
RT Drosophila white gene."
RL Gene 185:77-85(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=DBA/2;
RC MEDLINE=96359154; PubMed=8703120;
RA Savary S., Denizot F., Luciani M.-F., Mattei M.-G., Chimini G.;
RA "Molecular cloning of a mammalian ABC transporter homologous to
RA Drosophila white gene."
RT Mamm. Genome 7:673-676(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21092576; PubMed=11162488;
RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,
RA Assmann G., Cullen P.;
RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene."
RL Biochem. Biophys. Res. Commun. 280:121-131(2001).
RN [4]
RP INDUCTION, AND PROBABLE FUNCTION.
RX MEDLINE=20261604; PubMed=10799558;
RA Venkateswar A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,
RA Mangelsdorf D.J., Edwards P.A.;
RT "Human white/murine ABC8 mRNA levels are highly induced in
RT lipid-loaded macrophages. A transcriptional role for specific
RT oxysterols."
RL J. Biol. Chem. 275:14700-14707(2000).
RN [5]
RP REVIEW.
RX MEDLINE=21474438; PubMed=11590207;
RA Schmitz G., Langmann T., Heimerl S.;
RT "Role of ABCG1 and other ABCG family members in lipid metabolism."
RL J. Lipid Res. 42:1513-1520(2001).
CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is

```

```

CC an active component of the macrophage lipid export complex. Could
CC also be involved in intracellular lipid transport processes. The
CC role in cellular lipid homeostasis may not be limited to
CC macrophages.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN BRAIN, THYMUS, LUNG,
CC ADRENALS, SPLEEN AND PLACENTA. LITTLE OR NO EXPRESSION IN LIVER,
CC KIDNEY, HEART, MUSCLE OR TESTES.
CC -!- INDUCTION: Strongly induced in macrophage cell line RAW264.7
CC during cholesterol influx. Induction is mediated by the liver x
CC receptor/retinoid x receptor (LXR/RXR) pathway.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U34920; AAB47738.1; -
CC EMBL; Z48745; CAA88636.1; -
CC EMBL; AF323659; AAK27442.1; -
CC MGI; MGI:107704; Abcgl.
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR005284; Pigment_permease.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; PD000006; ABC_transportr; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRfams; TIGR00955; 3a01204; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 414
FT TRANSMEM 415 433
FT DOMAIN 434 444
FT TRANSMEM 445 465
FT DOMAIN 466 494
FT TRANSMEM 495 513
FT DOMAIN 514 521
FT TRANSMEM 522 543
FT DOMAIN 544 555
FT TRANSMEM 556 574
FT DOMAIN 575 637
FT TRANSMEM 638 657
FT DOMAIN 658 666
FT NP_BIND 118 125
SQ SEQUENCE 666 AA; 74033 MW; EDDC6AFBD43950B6 CRC64;

Query Match 70.3%; Score 2353; DB 1; Length 666;
Best Local Similarity 72.2%; Pred. No. 3.2e-150;
Matches 451; Conservative 80; Mismatches 90; Indels 4; Gaps 2;

Qy 26 DGAPPPVLTTHLKKVENHTEAQRFSHLPKRSAYDIEFVELSVYRGPCWRKRGYKTL 85
Db 42 DEVETDLLNGHLKVDNNFTFAQRFSPLPRAAVNIEFKDLSYVPEGPWKKYKTL 101
Qy 86 KCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRESGMQGLVNGRPRELTFKRMSCY 145
Db 102 KGISGKNSGELVAIMGPSGAGKSTLMNLAGYRETGMKGVAVLNGMRDLRCFRKVCY 161
Qy 146 IMQDDMLLPHTLVLEAMVSNANLNTENPDVKNLVTILTAGLMSCSHTRTALLSGGQ 205
Db 162 IMQDDMLLPHTLVQEAAMVSAHLQKDEGRREVMKIEITLALGLPCANTRTSLSGGQ 221
Qy 206 RKRLAIALELVNPPVFMFDEPTSGLDASCFQVVMKSLAQCGRITICTIHPSAKLF 265
Db 222 RKRLAIALELVNPPVFMFDEPTSGLDASCFQVVMKSLAQCGRISIVCTIHPSAKLF 281
Qy 266 EMFDKLYLSGGQCFKGVVNTNLPYKGLGHCPTVHNPAFTIEVASGEYGLNPMFL 325

```



```

WHIT_ANOGA
ID WHIT_ANOGA STANDARD; PRT; 695 AA.
AC Q27256; 017006;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Suakoko / G3;
RX MEDLINE=96423158; PubMed=8825759;
RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfler D.,
RA Collins F.H.;
RT "Cloning and characterization of the white gene from Anopheles
gambiae".
Insect Mol. Biol. 4:217-231(1995).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U29486; AAC46995.1; -
CC EMBL; U29485; AAC46994.1; -
CC EMBL; U29484; AAC47423.1; -
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR003439; ABC_transportr.
CC InterPro; IPR005284; Pigment_permease.
CC ProDom; PD000006; ABC_tran; 1.
CC SMART; SM00382; AAA; 1.
CC TIGRFAMS; TIGR00955; 3a01204; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 133 140 ATP (POTENTIAL).
FT NP_BIND 288 295 ATP (POTENTIAL).
FT TRANSMEM 444 464 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT TRANSMEM 524 544 POTENTIAL.
FT TRANSMEM 552 572 POTENTIAL.
FT TRANSMEM 581 601 POTENTIAL.
FT TRANSMEM 669 689 POTENTIAL.
FT CARBOHYD 472 472 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 100 100 N -> S (IN REF. 1; AAC47423).
FT CONFLICT 691 693 SRS -> YAR (IN REF. 1; AAC47423).
SQ SEQUENCE 695 AA; 77218 MW; EE8B9517239B2961 CRC64;

Query Match
Best Local Similarity 31.7%; Pred. No. 7.6e-45;
Matches 204; Conservative 117; Mismatches 276; Indels 46; Gaps 14;

QY 40 VENHITEAQRSHLPKRSAVDIEFE--LSYSVREGPCWRKR-----YKTLK 86
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 61 VKGIPOCERLTWTWKE--IDV-FGEAPTQDKPREPLCTRLRNCCTQRQKDFNPRKHLK 117
QY 87 CLSKKCFRRELIGMGPSGAGKSTFMILAGYRESGK-----GQILVNGRPRLRFRKM 142
D : : : : : : : : : : : : : : : : : : : : : : : : : :
D 118 NVTVGAKSGELLAVMGSSGAGKTTLLNALAFRSPPGVKISPNVAVNGVFNVAEQLRAR 177

RESULT 8
WHIT_ANOAL
ID WHIT_ANOAL STANDARD; PRT; 709 AA.
AC Q16928;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE White protein.
GN W.
OS Anopheles albimanus (New world malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC Culicoidae; Anopheles.
OX NCBI_TaxID=7167;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Teco;
RX MEDLINE=98126549; PubMed=9465401;
RA Ke Z., Benedict M.Q., Cornel A.J., Besansky N.J., Collins F.H.;
RT "The Anopheles albimanus white gene: molecular characterization of
the gene and a spontaneous white gene mutation.";
RL Genetica 101:87-96(1997).
CC -!- FUNCTION: MAY BE PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM
NECESSARY FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT
CELLS RESPONSIBLE FOR EYE COLOR.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC 143 SCYIMQDDMLLPHLTVLEAMMVSANLNTEN--PDVKNDLVTEILTALGLMSCSHT---- 196
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 178 CAYVQDDLFIPSLTTREHLLFQAMLRMGDRVPSAVKQHRVQEVLYQLSILVKCADTIIGA 237
QY 197 --RTALLSGGORKRLAIALELVNNPVMFFDEPTSGLDSSASCFOVSLKMSLAQGGRTII 254
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 238 PGRIKLSGGERRKLAFASETLDPHLLCDEPTSGLDSPMAISVLOVLKGMAMKGTII 297
QY 255 CTIHQPSAKILFEMPKDLYILSQQCIFKGVVNTLIPYLKGLGLHCPTYHNPADFIIE-- 311
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 298 LTIHOPSSSELYCLFDKILLVAEGRVAFGLSPYQSAEFFSOLGICPPNPYNPADFYVQMLA 357
QY 312 VASGEYGDLPNMLFRAVONGLC-AMAEKKSPEKNEVPAPC-----PPCPPEVDPTESH 364
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 358 IAPAKAEACRDMIKK-----ICDSFAVSPIAREVLLETASVAGKMGDEPYMLQQVEGVG 412
QY 365 TFSATSTUQF-CILFKRFTLSILRDTVLTLRPMFHVHVGVLGLLVLHIGDDASKVFNN 423
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 413 GYRSSWWTQFYCILW-RSWLSVLKDPMLVAVRLLQTAMVATLIGSVFGQVLDODGYVNI 471
QY 424 TGCLEFSSMLFMFAALMPTVLTTPLEMAVFMREHLNWTYSLKAYYLAKTMADVPFQVCP 483
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 472 NGSLFLFTNMTFONVFAVINVSAPLVELREKRSRLYRVDYVFLGKTTIAELPLFTAVP 531
QY 484 VVYCSIVYVMTGQPAETSRLLFSALATATATYALQAQSLGILLIGASNSLOVATFVGPVTAI 543
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 532 FVTSITYPMIGURTGAHYLTTLFIVTLVANVSTSGYLISCASSISWALSVPVVI 591
QY 544 PVLFSGFFYSFKTIPYQLQWSSVSVYRVYGFEGVILTYG-MERGLDTCLEERCPPREP 602
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 592 PELIFGGFELNSASVPAYFYKLSVSWFYANEALLINQWSTVVVDGEIACRANVTCPRS 651
QY 603 QSTLRALDVEDAKYLMDFVLGIFFLALRLAILVLRVYKSE 645
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
D 652 EIILETFENFRVEDFALDIACLFALIVLFRGLGALCLLWLRSRK 694

```

```
CC EMBL: L76302; AA88240.1; -.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF0005284; Pigment_permease.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR TIGRFAMs: TIGR00955; 3a01204; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 136 143
FT ATP (POTENTIAL).
FT NP_BIND 292 299
FT ATP (POTENTIAL).
FT TRANSMEM 457 475
FT POTENTIAL.
FT TRANSMEM 487 507
FT POTENTIAL.
FT TRANSMEM 537 555
FT POTENTIAL.
FT TRANSMEM 564 585
FT POTENTIAL.
FT TRANSMEM 598 616
FT POTENTIAL.
FT TRANSMEM 681 700
FT POTENTIAL.
FT CARBOHYD 485 485
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 658 658
FT N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE 709 AA; 79052 MW; FBD76D5C69D3BECA CRC64;

Query Match 22.0%; Score 736.5; DB 1; Length 709;
Best Local Similarity 32.4%; Pred. No. 5.4e-42;
Matches 199; Conservative 97; Mismatches 244; Indels 75; Gaps 13;

QY 82 KTLKCLSGKFCRRELIGIMSGAGKSTFMNLAGYRESGMK----GQILVNGRPRELR 137
DB 116 KHLKNVTGVARSGELLAVMGSSGAGKTTLLNELAFRSPGVKISPNAIRTLNGVPVTA 175
QY 138 TPRKSCYIMQDDMLPHLTIVLEAMVMSANLNTEN---PDVKNDLVTEILTALGLMSCS 194
DB 176 QMRARCAVQDDLFPSLTTKEHLMFQMLRGRDVPATPIKMRHVDEVLEQELSLVKA 235
QY 195 HT-----RTALLSGQRKRLAIALELVNPNPVMFFDEPTSGDASCFQVYSLMKSLAQ 248
DB 236 DFIIGVAGRVKGLSGERKRTAFRSETLTDPHLLCDEPTSLDSFMAQSVLOVLKGMAM 295
QY 249 GORTIICITHOPSAKLFEMFDKLYILSGQCQIFKGVVWNLIPYLGKGLGKLYTYPNPA 308
DB 296 KGKTIILTHOPSSSELYCLFDRILLVAEG-VAFGLSPYQSADFFSGLGICPCPNY 354
QY 309 -----IIIVASGEYGDNLNPNLFRVQNLCA 334
DB 355 YVQMLAIAPNKTETRETIKKICDSFVSPIDRIETASQVNGD-----GGIE 403
QY 335 MAEKKSSPEKNEVPAPCPPEVDPPIESHTSTATLTQF-CILFKRTLSILRDTVLTH 393
DB 404 LTRTKHTD-----PYFLQPMGEVDSTGYRASWWTQFYCILW-RSLVSLKDPMLVK 454
QY 394 LRFMSHVIVGLIGLLYLHIGDDASKVFNNTCGLFFSMLFLMFAALMPTVLTFLPMAYF 453
DB 455 VRLQOTAMVASLIGSIYFGQVLDQGVNMINGSLFLTNMTFQNVFAVINVESAEPLVF 514
QY 454 MREHLNLYSKAYLAKTMADVPQVVCPIVYVMTQPAETSRFLFLSALATAT 513
DB 515 LREKRSRLRVDTYFLGKTIABELPLFIAPVFTSITPYMIGLKAISHYLTFLTIVTLV 574
QY 514 ALVAQSLGLLGAANSLOVAFVGPVTAIPVLLSFGFEVSPKTIPTYLOWSSYLSYVRV 573
DB 575 ANVSTSFGLYLSCASSISMAISLVGPPVVPVIFLFGGFLNLSASVPAYFKYLSYLSWFRY 634
QY 574 GREGVILTYGNER-GDLTCLER--CPPREQSLRALD--VEDAKLYMDFLVIGIFPL 628
DB 635 ANEALLINOWADHRDGEIGCTRANVTCP-ASGEIILETFNFRVEDFALDIGCLFALIVLF 693
QY 629 ALRLLAYLVLYRVK 643
DB 694 RLGAFLCLMLRSRK 708

RESULT 9
YPC3_CAEEL
ID YPC3_CAEEL STANDARD; PRT; 598 AA.
```

```
AC Q1180;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative ABC transporter C05D10.3 in chromosome III.
GN C05D10.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL: U13645; AAA20989.2;
DR WormPep; C05D10.3; CE29170.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRFAMs: TIGR00955; 3a01204; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
KW Hypothetical protein; ATP-binding; Transmembrane; Transport.
FT NP_BIND 27 34
FT ATP (POTENTIAL).
FT TRANSMEM 336 356
FT POTENTIAL.
FT TRANSMEM 425 445
FT POTENTIAL.
FT TRANSMEM 453 473
FT POTENTIAL.
FT TRANSMEM 478 498
FT POTENTIAL.
SQ SEQUENCE 598 AA; 66906 MW; 9D6414E06898E343 CRC64;

Query Match 21.6%; Score 723.5; DB 1; Length 598;
Best Local Similarity 28.4%; Pred. No. 3.3e-41;
Matches 171; Conservative 126; Mismatches 255; Indels 51; Gaps 9;

QY 78 KRGYKTLKCLSGKFCRRELIGIMSGAGKSTFMNLAGYRESGM--KQOILVNGRPRE 135
DB 3 KRKRVIELHNVSQMAESGKLLALLGSSGAGKTTLMNVLRSNLTNLDVQSLIDGRAN 62
QY 136 LRTFRKMSYIMQDDMLPHLTIVLEAMVMSANLNTEN---PDVKNDL-VTEILTALGLMS 192
DB 63 KWKIREMSAFVQHQDMFVCTMTAREHLQFMARLMDGQYVSDHERQLRVEQVLTQMLKK 122
QY 193 CSHTRTAL-----LSGGQRKRLAIALELVNPNPVMFFDEPTSGDASCFQVYSLMKSL 246
DB 123 CADTVIGIPNQLKSGCEKRLSFASEILTCPILFCDEPTSGDLDFMAGHVQVLRSL 182
QY 247 AGCGRTIICITHOPSAKLFEMFDKLYILSGQCQIFKGVVWNLIPYLGKGLGKLYTYPV 306
DB 183 ADNGMTVIITIHQPSHHVYSLFNVCMLMACGRVIYILGPGDQVAPLFEKGYPCPAYNFA 242
QY 307 DFIE-----VASGEYGDNLNPNLFRVAV-----QNGICAMAEKSSPEKNE 346
DB 243 DHLIRTLAVIDSDRATSMKTIKIRQGFSLTDLQSVLAIGNANKLRAASFVTSGETSEK 302
QY 347 VPAPCPPEVDPIESHTFATSLTQFCILFKRTLSILRDTVLTHLRFMSHVIVGLI 406
```

Db 303 TKT-----FFNQYNASFWTQFLALEWRSLTVIROPNLLSVRLQLITAFIT 351
 QY 407 GLLYLHIGDASKVFNNTCCLFSLFALMPVTLFPLEMAYFMEHLNWTSLKA 466
 Db 352 GIVFFOTPVTPATISGIMFNHNRNMFMLOFPNVPVITAEPLIVLRENANGVYRTSA 411
 QY 467 YLIAKTMDVPFQVGVVYCVVYVYVMTGPAETSRFLFSALATATATALVAGSLGLIGA 526
 Db 412 YFLAKNIAELPOLYILPILYINIVYVMSGLYFNFNVCASLVTLITNVAISISAVAT 471
 QY 527 ANSLOVATFVGPTAIPVLLSGFVSKTPTIYLOWSSYLSYVRYGFEGLVITYGME 586
 Db 472 IFANTDVAITLPIFVVPVMAFGFEITDAIPSYFKWLSLSYFKYGEALAINEDWSI 531
 QY 587 RGDLTCLC-----ERCPEFPOSILRALDVEDAKLYMDLVGLGIFFLALRLAYLV 638
 Db 532 KVIPECNFSMTAFALDSCP-RNGHOVLESIDFSASHKIFDISILFGMFIGIRIAYVAL 590
 QY 639 RYR 641
 Db 591 LIR 593

RESULT 10
 ID ABG2_HUMAN STANDARD; PRT; 655 AA.
 AC Q9UN00; O95374; Q9UN00; Q9BY73;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE ATP-binding cassette, sub-family G, member 2 (Placenta-specific ATP-binding cassette transporter) (Breast cancer resistance protein).
 GN ABCG2 OR ABCP OR BCRP OR BCRP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=99065313; PubMed=9850061;
 RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;
 RT "A human placenta-specific ATP-binding cassette gene (ABCP) on chromosome 4q22 that is involved in multidrug resistance";
 RL Cancer Res. 58:5337-5339(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast cancer;
 RX MEDLINE=99080071; PubMed=9861027;
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RT "A multidrug resistance transporter from human MCF-7 breast cancer cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).
 RN [3]
 RP ERRATUM.
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,
 RX Ross D.D.;
 RL "Breast cancer resistance protein constitutes a 140-kDa complex as a homodimer";
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 198-655 FROM N.A.
 RC TISSUE=Placenta;
 RX ISOgai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RW Wagatsuma M., Hosofiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,

RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project";
 RN Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP REVIEW.
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Helmerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 CC -!- FUNCTION: XENOBOTIC TRANSPORTER THAT APPEARS TO PLAY A MAJOR ROLE
 IN THE MULTIDRUG RESISTANCE PHENOTYPE OF A SPECIFIC MCF-7 BREAST
 CANCER CELL LINE. WHEN OVEREXPRESSED, THE TRANSPORTED CELLS BECOME
 RESISTANT TO MITOXANTHONE, DAUNORUBICIN AND DOXORUBICIN, DISPLAY
 DIMINISHED INTRACELLULAR ACCUMULATION OF DAUNORUBICIN, AND
 MANIFEST AN ATP-DEPENDENT INCREASE IN THE EFFLUX OF RHODAMINE 123.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF103796; AAD09188.1; -;
 DR EMBL; AF098951; AAC97367.1; -;
 DR EMBL; AB056867; BAB39212.1; -;
 DR EMBL; AK002040; BAA92050.1; -;
 DR Genew; HGNC:74; ABCG2.
 DR MIM; 603756; -;
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transportr.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transportr; 1.
 DR SMART; SM00382; AAA; 1.
 DR PROSITE; PS00211; ABC_TRANSPORTER; FALSE_NEG.
 KW ATP-binding; Transmembrane; Transport.
 FT DOMAIN 1 395
 FT TRANSMEM 396 416
 FT DOMAIN 417 428
 FT TRANSMEM 429 449
 FT DOMAIN 450 477
 FT TRANSMEM 478 498
 FT DOMAIN 499 506
 FT TRANSMEM 507 527
 FT DOMAIN 528 535
 FT TRANSMEM 536 556
 FT DOMAIN 557 630
 FT TRANSMEM 631 651
 FT DOMAIN 652 655
 FT NP_BIND 80 87
 FT CARBOHYD 418 418
 FT CARBOHYD 557 557
 FT CARBOHYD 596 596
 FT CONFLICT 24 24
 FT CONFLICT 166 166
 FT CONFLICT 208 208
 FT CONFLICT 315 316
 FT CONFLICT 482 482
 FT CONFLICT 555 555
 FT CONFLICT 655 655
 FT SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CCE0 CRC64;
 SQ
 Query Match 21.3%; Score 712; DB 1; Length 655;
 Best Local Similarity 27.8%; Pred No. 2 2e-40;
 Matches 177; Conservative 141; Mismatches 246; Indels 72; Gaps 16;
 QY 61 IEFVELSVSV--REG--PCWRRKRYKYLKCLSGKRCRELIGIMGPGAGKSTFMNLA 116
 Db 37 LSFHNICRYVKLKGSLPC-RKPVEKEILSNING-IMKPLGNAILGPTGGKSSLDVLA 94

```

QY 117 GYRE-SGKGOILVNGRPRELRTFRKMSCYIMQDDMLLPHLTIVLEAMVMVSANLNL--TEN 173
DB 95 ARKDPGSLGDLVINGAPRP-ANEKCNKSGYVQVQDDVMTLVRENLOFSAALRLATTMT 153
QY 174 PDVKNLVTLLTALGLMSCSHRTAL-----LSGQGRKRLAIALELVNPPVVFDEPT 228
DB 154 NHEKRNIRNVEELGDKVADSKVGTQFIRGVSGGKRRTSIGMEITDPSILFLDEPT 213
QY 229 SGLDSASCFQVSLMKSIAOGRRTIICTHOPSASAKLFEMFKLILSGOGCIFKGVVTNL 288
DB 214 TGLDSSANAVLLLLKMSKGRRIIFSIHOPRYSIEKLFDSLILLASGRMLFHPQAE 273
QY 289 IPYLKGLGHCHPTVHNPAFTIEVASGEYDNLPMFLFRAVONGICLAMAEEKSSPEKNEVP 348
DB 274 LGYFESAGYHCEAYNNPAFDLDINGD-----STAVALNREEDFKATEII 319
QY 349 APCPPCEVDPI-----ESH-----TFATSLTQFCI 376
DB 320 EPSKODPLIKLAEIYVNSFYKETRAELHQLSGGKKKIVFKETISYTTSCHEQLRW 379
QY 377 LFKRTFSLRDTVLTLRFMSHVVIGVLIGLLYHIGDDASKVFNNTGCLFFSMLFLMF 436
DB 380 VSKESFNLLGNPOASTAIIVTVGLVIGAIYGLKNDSTGIONRAGVLFLLTNOCP 439
QY 437 AALMPTVLTPLPMVAFMREHLNWSLYKAYIAKTMAV-VPFQVGPVVCYIVYMWGT 495
DB 440 SSVSAVEL-FVYERKLETHEYISGYRYVSSYFLGKLLDPLMRMLPSIIFTCIVYFMLG 498
QY 496 QPAETSRELLFSALATATLVAQSLGILLIGASNSLOVATFVGPVTAIPVLLFSGFFVSP 555
DB 499 LKPADAFFVMMFTLMVAVSASSMALAIAAGQSVSVATLLMTICFVFMFISGLLVNL 558
QY 556 KTIPTYLQWSSYLVRYGFGVITIV-----GME-RGDLTCLERCPREPOQILR 607
DB 559 TTIASLWSLQYFISIPRYGFTALQHNELFQNFQGLNATGNNVCNATCI-GRYLVKQ 617
QY 608 ALVEDAKLVNDFVLVGLIFFLALRLAYLVLYRVK 643
DB 618 GIDLSPWGLWKNHVALACMIVIFLTIAVYKLLFLKK 653

RESULT 11
SCRT_DROME
ID SCRT_DROME STANDARD; PRT; 666 AA.
AC P45843;
01-NOV-1995 (Rel. 32, Created)
01-FEB-1996 (Rel. 33, Last sequence update)
16-OCT-2001 (Rel. 40, Last annotation update)
DE Scarlet protein.
GN ST.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RA Garcia R.L., Perkins H.D., Howells A.J.;
RL Submitted (OCT-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 200-306 FROM N.A.
RX MEDLINE=89339145; PubMed=2503416;
RA Tearle R.G., Belote J.M., Mckown M., Baker B.S., Howells A.J.;
RT "Cloning and characterization of the scarlet gene of Drosophila
RT melanogaster.";
RL Genetics 122:595-606(1989).
CC -!- FUNCTION: PART OF A MEMBRANE-SPANNING PERMEASE SYSTEM NECESSARY
CC FOR THE TRANSPORT OF PIGMENT PRECURSORS INTO PIGMENT CELLS
CC RESPONSIBLE FOR EYE COLOR. SCARLET AND WHITE DIMERIZE FOR THE
CC TRANSPORT OF TRYPTOPHAN.
CC -!- SUBUNIT: HETERODIMER OF SCARLET WITH WHITE.

```

```

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39739; AAA82056.1; -
DR EMBL; X76201; CAA53794.1; -
DR FlyBase; FBgn0003515; st.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR005284; Pigment_permease.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfams; TIGR00955; 3a01204; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW Pigment; ATP-binding; Transmembrane; Transport.
FT NP_BIND 108 115 ATP (POTENTIAL).
FT TRANSMEM 418 438 POTENTIAL.
FT TRANSMEM 445 465 POTENTIAL.
FT TRANSMEM 491 511 POTENTIAL.
FT TRANSMEM 519 539 POTENTIAL.
FT TRANSMEM 552 572 POTENTIAL.
FT TRANSMEM 577 597 POTENTIAL.
FT TRANSMEM 640 660 POTENTIAL.
SQ SEQUENCE 666 AA; 74506 MW; 6796ED4084B59CE4 CRC64;

Query Match 19.9%; Score 667; DB 1; Length 666;
Best Local Similarity 29.0%; Pred. No. 2,3e-37;
Matches 186; Conservative 128; Mismatches 258; Indels 70; Gaps 20;

QY 53 LPKRSVDIEFVELSYS-----VREGP--CWKR-----RGYTKLLKCLSGKFCRR 95
DB 42 LSKRNSERSLPLRSYSKWSPTQCATLVWRDLCVTVTVGGSGQRMKRIINNSTCAIQPG 101
QY 96 ELIGMGPSGAGKSTFNILACYRESG--MKGOIIVNGRPRELRTFRKMSCYIMQDDMLL 153
DB 102 TLMALMGSSGSKTTLMTSLAFQPAQTVVQGDLLNGR-RIGPFMRHNGHYVYQDDLLFL 160
QY 154 PHLTVLEAMVVSANLNLTE--NPQVKNLVTETLTLGLMSCSHRTA-----LLSGQ 205
DB 161 GSVSVLEHLNFMMAHLRLDRRVSKERRLIIEKELLERTGLLSAAQTRIGSGDDKKVLSGGE 220
QY 206 RKRLAIALELVNPPVVFDEPTSGLDSASCFQVSLMKSIAOGRRTIICTHOPSASAKLF 265
DB 221 RKRLAFAVELNPNVILFCDEPTTGLDSYSAQQLVATLYELAQAQGTTLTCTIHOPSSQLF 280
QY 266 EMFKLYILSGOCIFKGVVTNLIPYLKGLGHCHPTVHNPAFTIEVASGEYDNLPMFLF 325
DB 281 DNFNNMILLADGRVAFVTGSPQHALSFFANHHGYCYEAYNPADFLIGLVATDPG-YEQASQ 339
QY 326 RAVONGICLAMAEEKSSPEKNEVPAPCPPEVDPIESTHTATST-----LTQFC 375
DB 340 RSAQH-LCDQFAVSSAAKQDMLVNL-----EIHMAQSGNFPFDTEVESFRGVAMVKRPH 393
QY 376 ILFKRTFLSLRDTVLTLRFMSHVVIGVLIGLLYHIGDDASKVFNNTGCLFF 429
DB 394 VVMLRAIVTLRDPITQWLRFIQIAMAFIIGACFAGTTPESQLGQVAQV-----GALFI 448
QY 430 SMLFLMFAALMPTVLTFFPLEMAVMREHLNWSLYKAYIAKTMAVDFQVQVVCYVCS 489
DB 449 MISENTYHPMYSVLNLFEPQGFLEFMRSTRSLYSTGYAANILALLPGMIEPLIFVII 508
QY 490 VYWMTGQPAETSRELLFSALATATL-----VAQSLGILLIGASNSLOVA-TFVGPVTAIPV 545
DB 509 CYWLTGL---RSTFYAFGVATWCVVVMNVNATACGCFSTAFNSVPLAMAYLVLPDIYI-F 564

```

	ENBL; X59720; CAA42328.1; .	
DR	PIR; S19421; S19421.	
DR	PIR; S40914; S40914.	
DR	SGD; S0000604; ADP1.	
DR	InterPro; IPR003593; AAA_ATPase.	
DR	InterPro; IPR003439; ABC_transportr.	
DR	Pfam; PF00005; ABC_tran; 1.	
DR	Frodcom; PD0000006; ABC_transportr; 1.	
DR	SMART; SM00382; AAA; 1.	
DR	PROSITE; PS00211; ABC_TRANSPORTER; 1.	
KW	ATP-binding; Transmembrane; Glycoprotein; Signal.	
FT	SIGNAL	1 25 POTENTIAL.
FT	CHAIN	26 1049 PROBABLE ATP-DEPENDENT PERMEASE.
FT	NP_BIND	423 430 ATP (BY SIMILARITY).
FT	TRANSMEM	325 345 POTENTIAL.
FT	TRANSMEM	464 481 POTENTIAL.
FT	TRANSMEM	794 814 POTENTIAL.
FT	TRANSMEM	829 849 POTENTIAL.
FT	TRANSMEM	878 898 POTENTIAL.
FT	TRANSMEM	910 930 POTENTIAL.
FT	TRANSMEM	938 958 POTENTIAL.
FT	TRANSMEM	1001 1021 POTENTIAL.
FT	TRANSMEM	1025 1045 POTENTIAL.
FT	CARBHYD	50 50 N-LINKED (GLCNAC...) (POTENTIAL).

```

RESULT 13
ABG5_HUMAN
ID ABG5_HUMAN          STANDARD;          PRT;      651 AA.
AC Q9H222;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).
GN ABG5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT GLU-604.
 RC TISSUE=Liver;
 RX MEDLINE=20553648; PubMed=11099417;
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,
 RA Kvitrovich P., Shan B., Barnes R., Hobbs H.H.;
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by
 mutations in adjacent ABC transporters.";
 RL Science 290:1771-1775(2000).
 RN [2]
 RP SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA H-389; H-419 AND P-419,
 RP AND VARIANT E-604.
 RC TISSUE=Liver;
 RX MEDLINE=20578753; PubMed=11138003;
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,
 Dean M., Patel S.B.;
 RT "Identification of a gene, ABCG5, important in the regulation of
 dietary cholesterol absorption.";
 RL Nat. Genet. 27:79-83(2001).
 RN [3]
 RP REVIEW
 RX MEDLINE=21474438; PubMed=11590207;
 RA Schmitz G., Langmann T., Heimerl S.;
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";
 RL J. Lipid Res. 42:1513-1520(2001).
 RN [4]
 RP VARIANTS SITOSTEROLEMIA Q-146; H-389; P-419; H-419 AND S-550, AND
 RP VARIANT E-604.
 RX MEDLINE=21344600; PubMed=11452359;
 RA Lu K., Lee M.-H., Hazard S., Brooks-Willson A., Hidaka H., Kojima H.,
 RA Ose L., Stallenhoef A.F.H., Miettinen T., Bjorkhem I., Bruckert E.,
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
 Patel S.B.;
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic
 structure and spectrum of mutations involving sterolin-1 and
 sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
 RL Am. J. Hum. Genet. 69:278-290(2001).
 CC -I- FUNCTION: Transporter that appears to play an indispensable role
 CC in the selective transport of the dietary cholesterol in and out
 CC of the enterocytes and in the selective sterol excretion by the
 CC liver into bile.
 CC -I- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to
 CC ABCG8 along a pathway regulating dietary-sterol absorption and
 CC excretion.
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -I- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels
 CC in the small intestine and colon.
 CC -I- DISEASE: Defects in ABCG5 are a cause of sitosterolemia, also
 CC known as phytosterolemia or shellfish sterolemia, a rare autosomal
 CC recessive disorder characterized by increased intestinal
 CC absorption of all sterols including cholesterol, plant and
 CC shellfish sterols, and decreased biliary excretion of dietary
 CC sterols into bile. Sitosterolemia patients have
 CC hypercholesterolemia, very high levels of plant sterols in the
 CC plasma, and frequently develop tendon and tuberosus xanthomas,
 CC accelerated atherosclerosis and premature coronary artery disease.
 CC -I- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCG (WHITE)
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF320293; AAC40003.1;
 CC EMBL: AF312715; AAC53099.1;
 CC Genew: HGNC:13886; ABCG5.

DR MIM: 605459; -
 DR MIM: 210250; -
 DR InterPro: IPR003593; AAA_ATPase.
 DR InterPro: IPR003439; ABC_transportr.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transportr; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; FALSE_NEG.
 KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism;
 KW Disease mutation.
 FT DOMAIN 1 383 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 384 404 1 (POTENTIAL).
 FT DOMAIN 405 421 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 422 442 2 (POTENTIAL).
 FT DOMAIN 443 462 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 463 483 3 (POTENTIAL).
 FT DOMAIN 484 503 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 504 524 4 (POTENTIAL).
 FT DOMAIN 525 528 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 529 549 5 (POTENTIAL).
 FT DOMAIN 550 623 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 624 644 6 (POTENTIAL).
 FT DOMAIN 645 651 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 86 93 ATP (POTENTIAL).
 FT CARBOHYD 584 584 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 146 146 E -> Q (IN SITOSTEROLEMIA).
 FT VARIANT 389 389 R -> H (IN SITOSTEROLEMIA).
 FT VARIANT 419 419 R -> H (IN SITOSTEROLEMIA).
 FT VARIANT 419 419 R -> P (IN SITOSTEROLEMIA).
 FT VARIANT 419 419 R -> S (IN SITOSTEROLEMIA).
 FT VARIANT 550 550 Q -> E.
 FT VARIANT 604 604 /FTID=VAR_012248.
 FT SEQUENCE 651 AA; 72503 MW; 950BABFCB6A1536 CRC64;
 Query Match 17.5%; Score 585.5; DB 1; Length 651;
 Best Local Similarity 27.4%; Pred. No. 6.3e-32;
 Matches 172; Conservative 125; Mismatches 277; Indels 53; Gaps 14;
 QY 54 PKRSADVIEFVLSYSVREGPCW-----RRKGYKTLKCLSGKFCRRELIGIMGPSGAGK 108
 DB 33 PEPHSLGILHASYSVSHRVRPMDITSCROQWTFQILKDVLSYVESQIMCLGSSGSK 92
 QY 109 STFMNLIAGY--RESGKKGOILVNGRPRLRTRFKMVCYIMQDDMLPLHLTVLEAMMVA 166
 DB 93 TTLLDAMSGRLGRAGTFLGEVYVNGRALRRREQQDCFSYVLQSDTLSSLTVRRETHYTA 152
 QY 167 NLNLTE-NPDVKNDLVTEIITALTGLMSCSHTRTAL-----LSGQKRLALALELVN 217
 DB 153 LLAIKRGNGPSFKKVEANMAELSL---SHVAORLIGNYSGLGISTGERRVSTAAQLQ 209
 QY 218 NPVPMFEDEPTSGIDSASCFQVYVSLKSLAQGGRTIITIHQPSAKLFEMFDKLYILSQG 277
 DB 210 DKPVMLFDEPTGLDCNTAQIVLVLLVARRNRIVVLTIHQPSSEFQFQDKTAILSFG 269
 QY 278 QCIFKGVVWNLIPYKGLGLHCHPTYNHPADFIIEVASG-----EYGDNLNMLPRAV 328
 DB 270 ELIFCGTPAEMDLFFNDCGYPCEHSNPFDFMDLTSVDTQSKREKETSKRVQMIESAY 329
 QY 329 -QNGLCAMAEKSKSPEKEVPAFCPPCPPEVDPIESHTFATSTLTQFCILFKRTLSILR 387
 DB 330 KKAICHKTKTLKNIERNMKHLKTLPMVPFKTKDSP-----GVFSKGLVLLRRVTRNLVR 381
 QY 388 DTVLTHLRKSHVVIGVIGLGLYLIHIGDDASK--VFNNTCGLFSPMLFMAALMPTVLT 445
 DB 382 NKLAVITRLLQNLIMGLFLFFVLRVRSNVLKGAIQDRVGLLQYFGVCAFTYTGMLNAYNL 441
 QY 446 FPLEMAVFMREHLNWNYSKAYYLAKTMADVPOFVQCPVYVCYSIVYVMTQGPATRSFL 505

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:38:29 ; Search time 44 Seconds
(without alignments)
1411.428 Million cell updates/sec

Title: US-10-072-621-9
Perfect score: 3347
Sequence: 1 MAEKALEAVGCGPGAVAM.....FLALRLAYLVLRVKSER 646
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

rched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 283224

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3309	98.9	646	2 JC7777	ATP binding cassette
2	2350	70.2	638	2 G02068	white homolog - hu
3	2161	64.6	416	2 T46401	hypothetical prote
4	879.5	26.3	687	1 FYFW	white protein - fr
5	760.5	22.7	646	2 C86441	probable ABC trans
6	716.5	21.4	658	2 T31958	hypothetical prote
7	699.5	20.9	609	2 E96742	probable ABC trans
8	697	20.8	737	2 T46101	ABC transporter-li
9	686	20.5	695	2 T21109	hypothetical prote
10	685	20.5	559	2 B88474	protein C05D10.3 l
11	665.5	19.9	725	2 C84423	probable ABC trans
12	663.5	19.8	590	2 B96573	protein F12M16.17
13	640	19.1	635	2 T08934	hypothetical prote
14	635.5	19.0	608	2 T34391	hypothetical prote
15	626	18.7	577	2 T04229	ABC-type transport
16	613	18.3	649	2 A84509	probable ABC trans
17	612	18.3	725	2 T47852	ABC transporter-li
18	611	18.3	678	2 H96552	hypothetical prote
19	607	18.1	633	2 T19189	hypothetical prote
20	600	17.9	639	2 G88939	protein C10C6.5 [i
21	598.5	17.9	720	2 T47648	ABC transporter-li
22	596	17.8	687	2 D96553	hypothetical prote
23	595	17.8	1049	1 S19421	ATP-dependent perm
24	586.5	17.5	739	2 T45891	ABC transporter-li
25	585.5	17.5	547	2 T31543	hypothetical prote
26	585	17.5	740	1 T02567	probable Amp-bindi
27	584.5	17.5	1435	2 D96693	protein Putative A
28	584	17.4	708	2 T47650	ABC transporter-li
29	577	17.2	610	2 T19333	hypothetical prote

30	573.5	17.1	755	2 G84791	probable ABC trans
31	571	17.1	1423	2 A86289	probable ABC trans
32	565	16.9	1294	2 S77690	probable membrane
33	563	16.8	1443	2 T02491	probable ABC trans
34	559	16.7	1450	2 A84780	probable ABC trans
35	555.5	16.6	1420	2 T02644	ABC-type transport
36	546.5	16.3	662	2 T47649	ABC transporter-li
37	545.5	16.3	659	2 E86313	hypothetical prote
38	535	16.0	1451	2 B86286	F9L1.15 protein -
39	534	16.0	1450	2 T45888	ABC transporter-li
40	533.5	15.9	1459	2 H96622	probable ABC trans
41	515	15.4	1413	2 G84790	probable ABC trans
42	486	14.5	1426	2 T30567	ATP-binding casses
43	483	14.4	1333	2 S63403	probable membrane
44	473	14.1	979	2 G85167	ABC transporter li
45	472	14.1	1529	2 S69688	hypothetical prote

ALIGNMENTS

RESULT 1
JC7777
ATP binding cassette transporter, subclass G4 - human
C:Species: Homo sapiens (man)
C:Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 02-Apr-2002
C:Accession: JC7777
R:Engel, T.; Lorkowski, S.; Lueken, A.; Rust, S.; Schlueter, B.; Berger, G.; Cullen, Biochem. Biophys. Res. Commun. 288, 483-488, 2001
A:Title: The human ABCG4 gene is regulated by oxysterols and retinoids in monocyte-de
A:Reference number: JC7777
A:Contents: Macrophages
A:Accession: JC7777
A:Molecule type: mRNA
A:Residues: 1-646 <ENG>
A:Cross-references: GB:AJ308237
C:Comment: This protein, a novel member of the ATP binding cassette(ABC) half-size tr
C:Genetics:
A:Gene: abcg4
A:Map position: 11q23.3
C:Keywords: macrophage

Query Match 98.9%; Score 3309; DB 2; Length 646;
Best Local Similarity 98.9%; Pred. No. 5.2e-251;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY	1	MAEKALEAVGCGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSADV	60
DB	1	MAEKALEAVGCGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSADV	60
QY	61	IEFVELSYSVREGPCWKRKGYKTLKCLSGKFCRRRELIGIMGPGACKSTFMNIIAGYRE	120
DB	61	IEFVELSYSVREGPCWKRKGYKTLKCLSGKFCRRRELIGIMGPGACKSTFMNIIAGYRE	120
QY	121	SGMKGQTLVNGRPRELTFKMSCYIMQDDMLLPHLTLEAMVSNANLNTENPDVKNL	180
DB	121	SGMKGQTLVNGRPRELTFKMSCYIMQDDMLLPHLTLEAMVSNANLNTENPDVKNL	180
QY	181	VTEILTALGLMSCSHRTALLSGGQRRLAIAELVNNPPVMPFDEPSTGLDSASCFQV	240
DB	181	VTEILTALGLMSCSHRTALLSGGQRRLAIAELVNNPPVMPFDEPSTGLDSASCFQV	240
QY	241	SLMKSLAQQGRTTICTIHQPSAKLFEMFDKLYILSQQCIFKGVVWNLIPYLKGLGHCP	300
DB	241	SLMKSLAQQGRTTICTIHQPSAKLFEMFDKLYILSQQCIFKGVVWNLIPYLKGLGHCP	300
QY	301	TYHNPAADFTIEVASGEYGLNPNMLFRAVQNGLCMAEKSSPEKNEVPAPCPPEVDP	360
DB	301	TYHNPAADFTIEVASGEYGLNPNMLFRAVQNGLCMAEKSSPEKNEVPAPCPPEVDP	360
QY	361	IESHTATSTLTQFCILFKRTFLSILRDTVTLHRLFMSHVIVGLTGLLYLHIGDASKV	420
DB	361	IESHTATSTLTQFCILFKRTFLSILRDTVTLHRLFMSHVIVGLTGLLYLHIGDASKV	420

R:Pepling, M.; Mount, S.M.
Nucleic Acids Res. 18, 1633, 1990
A:Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.
A:Reference number: S08635; MUID:90221897; PMID:2109311
A:Accession: S08635
A:Molecule type: mRNA
A:Residues: 1-687 <PEP>
A:Cross-references: EMBL:X51749; NID:g8825; PIDN:CAA36038.1; PID:g8826
R:O'Hare, K.; Murphy, C.; Levis, R.; Rubin, G.M.
J. Mol. Biol. 180, 437-455, 1984
A:Title: DNA sequence of the white locus of *Drosophila melanogaster*.
A:Reference number: S07263; MUID:85134865; PMID:6084717
A:Accession: S07263
A:Molecule type: DNA
A:Residues: 1-24, 'LIFEIPYHCRVTAD', 30-334, 'ITLHNSYPAPWPSVLPTTIRRTFYRCWPLCPDGRSSPVIGSPR'
A:Cross-references: EMBL:X02974
A:Experimental source: strain Canton S
O'Hare, K.
Submitted to the EMBL Data Library, June 1985
Reference number: S10240
A:Accession: S10240
A:Molecule type: DNA
A:Residues: 1-24, 'LIFEIPYHCRVTAD', 30-687 <OHA2>
A:Cross-references: EMBL:X02974; NID:g10873; PIDN:CAA26716.1; PID:g10874
A:Experimental source: strain Canton S
C:Genetics:
A:Gene: white; w
A:Cross-references: FlyBase:FBgn0003996
A:Introns: 24/3; 116/1; 334/2; 439/3; 483/3
A:Superfamily: fruit fly white protein; ATP-binding cassette homology
C:Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein
F:113-317/Domain: ATP-binding cassette homology <ABC>
F:130-137/Region: nucleotide-binding motif A (P-loop)
F:261-265/Region: nucleotide-binding motif B
F:67, 93, 472, 554, 651/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 26.3%; Score 879.5; DB 1; Length 687;
Best Local Similarity 33.9%; Pred. No. 9.6e-61;
Matches 206; Conservative 121; Mismatches 237; Indels 43; Gaps 12;
QY 73 GPCWRK-----RCY-----KTLKCLSGKCRRELIGIMPSGAGKSTFNNILA 116
DB 85 GSGWRQLVNTRGFLNCRHPIAPRHLKKNVCGVAYPGELLAVMGSSGAGKTKTLNALA 144
QY 117 GYRESGMK-----GQILVNGRPRLRTFRKMSCYIMQDDMLLPHLTVEAMVMSANLNLT 172
145 FRSPQGIQVSPSGMRLNGQVDAKEMQARCAVYQDDFLIGSLTAREHLIFQAVRMPR 204
QY 173 NPDKNDL--VTEILTALGLMSCHT-----RTALLSGQKRLAIALELVNPPVWF 224
DB 205 HLTYRQVARVDQVQIELSLKCOHTIIGVGRVKGSLGGERKRLAFASEALDPPLLIC 264
QY 225 DEPTSGLDASCFQVYSLMKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGCICFKV 284
DB 265 DEPTSGLDSTANUSVQVQLKLSQKGTVLTTHQPSSELFDFKILLMAEGRVAFGLT 324
QY 285 VTNLIPYKGLGHCPTHNPADFIIEVAS--GEYGDLPMLFRAVQNGLCAMAEKSS 341
DB 325 PSEAVDFSVGACQPTNYPADFYQVLAVPGREIESDRKACIDN----FATSKVA 380
QY 342 PERNEVPAPCPPEVDPTESHTFATSTLTQCIILKRTFLSILRDTVLTHLRFMSHV 401
DB 381 RDMEQLLATKNLEKLEQPENGYTYKATPMQFRAVLWSLSVLEKLEPLLVKRLIQT 440
QY 402 IGVILGLYLLHIGDASK--VFNTGCLFFSMFLMFAALMPVLTLPPELVAFMVRHLN 459
DB 441 VAILIGLIFL--GQULTQVGMNNGAIFLFLNMTFQNVFATINVTSELPMRARS 498
QY 460 YWISLAKYLAKTMAVDFQVCPVYVCSIVYWMTPGAPETSRELLFSALATATVAQS 519
DB 499 RLYRCDTYFLGKTIAELPLELTPLVFTATAYPMIGLRAGVLHFFNCALVTLVANV 558
QY 520 LGLLGAANSLOVATFVGVPVTAIPVLLSGFFVSKFTIPTYLQWSSYLSVRYGEGVI 579

DB 559 FGYLISCASSSTSMALSVGPPVPIPELLEGGFFLNSGSPVYKWLKLSYLSWERYANEGL 618
QY 580 LTIYG-MERGDLTCLER--CPREPOSILRALDVEDAKLYMDFLVIGLFFLRLAYL 636
DB 619 INQWADVEPGEISCTSSNTTCC--SSGKVLLETUNFSAADPLDYGLAILIVSRVLA 677
QY 637 VLRYRVK 643
DB 678 ALRLRAR 684
RESULT 5
C86441
probable ABC transporter [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86441
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 815-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-646 <STO>
A:Cross-references: GB:AE005172; NID:g1136734; PIDN:AAG31315.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F1216.1; AT

Query Match 22.7%; Score 760.5; DB 2; Length 646;
Best Local Similarity 30.3%; Pred. No. 1.9e-51;
Matches 189; Conservative 133; Mismatches 239; Indels 63; Gaps 19;
QY 57 SAYDIEFVELSVSVR---EGPC---WRKGYKTLKCLSGKCFRRELIGIMPSGAGKST 110
DB 47 SMYPITLKEVYKVKTEQTSQCMGSKK-EKTLNGITGWCPCGEFLAMLGPSGSKT 105
QY 111 FMILAGYRESGMKQILVNGRPRLRTFRKMSCYIMQDDMLLPHLTVEAMVMSANLN 170
DB 106 LLSALGGRUSKTFSGVMYNGQPFSGCTKRRTGFAQDDVLPHTLVETLFTALLRL 164
QY 171 TEN--PDVKNLDVTEIITLALGMSCHTRTA----LLSGQKRLAIALELVNPPV 223
DB 165 PSSLTRDEKAHVDVRIAEGLNRCNSMIGGLPFRGISGGEKKRYSIGEMINSL 224
QY 224 FDEPTSGLDASCFQVYSLMKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGCICFK 283
DB 225 LDEPTSGLDSTTAHRTVITIKRLASGRTVVTTIHQPSRIYHMFQKVLSESPY 284
QY 284 VTNLIPYKGLGHCPTHNPADFIIEVAS--GEYGDLPMLFRAVQNGLCAMAEK 339
DB 285 AASAEYFESSLGFTSLAVNPADLLDLANGIPDPTQKETSEQEQTKVETLVSAYEK 344
QY 340 SSPEKNEVPAPCPPEVDPTESHTF----ATSTL--TQCFILKRTFLSILRDTV 390
DB 345 ISTK-----LKAELCNAESHSEYTKAAKNLKSQWCTTWYQVTLVLRQVRE 394
QY 391 -----LTHLRFMSHVIGVLIGLHYLHIGDDASKVFNNTCCLFFSLMFLMFAALMPTVLT 445
DB 395 RPESEFNKLRIFQVISVAFGLGLLWH--TPKSHIQDRTALLFFSVFVGFYPLVNAVFT 452
QY 446 FPLEMAVFMREHLNHYWLSKAYLAKTMADVPPQVCPVYVCSIVYWMTPGAPETSRELL 505
DB 453 FPOEKRMLIKERSGMYRLSSYFMARNVGDPLELALPTAFVFIYWMGGLKPDPTFTIL 512

484	Db	TILPMYGTIIIIYAGLVASVTSLSVLFVFCITLTWVAVSIA-YVGACIFGDEGLVVTM	542
538	QY	GPVTAIPVLLFSGFVFSFKTIPTYLQNSYLSVRYRGEGV-----ILTIYGMRGDL	590
543	Db	-PMEVPLMLVFGGVVANSIPVYQYVSVFSFKHGFEALEANQWKDKISGC---DL	598
591	QY	-----TCLLEERCPEFQSIIRALDVEDAKLYMDFLVGLFFLIALLRLAVLVURYVK	643
599	Db	INPLNATTGVCPSADGPIILTRGI--DTPLYANVILTPMSFVYKRIIGLVALKIKRV	655

RESULT 7
E96742
probable ABC transporter FLTM19.11 [imported] - Arabidopsis thaliana
C-species: Arabidopsis thaliana (mouse-ear cress)
C-date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C-accession: E96742
R:R.Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; I.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.;
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E96742

[illegible]

Db 428 FMAHVLGSLSMELVLPASFLTFTYWMYVLRPGIVPELTLTSLVLLLYVLASQGLGALGAA 487
QY 528 SNSLOVATFVGPVTAIPVLLSGFEVFSFKTIPTIYLOWSSYLS- - - - -YVRGFGF 577
Db 488 IMDAKKASIVIVTMFLAVLTGYYVN--KPSGWMWYKYSTTFYCYRLVAIQSGSE 545
QY 578 VILTYGME- - - - -RGDLTCLERCPCFPREPOSILRALDVEDAKLYMDFLVGLGIFFLALRL 633
Db 546 EILRLMGCDCKQKQAGSAATSAGCRFVEEVI- - - - -GDVGMWTSVGVLFLEMFYRYL 599
QY 634 AYLVLRVRYK 643
Db 600 AYLAIR-RIK 608

RESULT 8

T46101
C: transporter-like protein - Arabidopsis thaliana
Alternate names: protein T25B15.80
Species: Arabidopsis thaliana (mouse-ear cross)
C: Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C: Accession: T46101
R: Alcaraz, J.P.; Clabault, G.; Cottet, A.; Mache, R.; Meves, H.W.; Lemcke, K.; Mayer, K.
submitted to the Protein Sequence Database, January 2000
A: Reference number: Z23021
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-737 <ALC>
A: Cross-references: EMBL:AL132972
A: Experimental source: cultivar Columbia; BAC clone T25B15
C: Genetics:
A: Map position: 3
A: Introns: 122/1; 146/3; 225/2; 277/2; 338/3; 422/2; 535/1; 628/3; 664/3
A: Note: T25B15.80

Query Match 20.8%; Score 697; DB 2; Length 737;
Best Local Similarity 28.5%; Pred. No. 2.1e-46;
Matches 187; Conservative 126; Mismatches 250; Indels 92; Gaps 20;
QY 37 LKKVENHITAEQRFSLPKRSADVIEFVLSYV-REGPCWRKRGYKTLKLSGKFCRR 95
Db 122 LEDIEAATSVVKFOAEP-TFFPYLKFIDITYKVTGK--MTSSSEKSLNGISGAYPG 178
QY 96 ELIGMPSGAGKSTPMNLAG-YRSGMKGOILVNGRP--RELRTFRKMSCYIMODDML 152
Db 179 ELLALMGPSGKTKTLNLAGRFGNQNGIGSVYNDKPYSKLKT---RIGFVTQDDVL 235
QY 153 LPHLTVLEAMVYSANLNL--TENPDVKNLVTLEILFALGMSCSHTR- - - - -TALLSGGO 205
Db 236 FPHLTVKETLTALTALLRLEKTLTEQKEQRAASVIGELGLERQDQTMIGGSFVRVSGGE 295
QY 206 RRLALALELVNPPVPMFDEPTSGLDASCFQVVSMLKSLAOGGRTICTTHOPSAKLF 265
Db 296 RKRVCIGNEMTNPSLLLDDEPTSSUDSTALKIVQMLHCIAKAGKTIYTTTHOPSSRLF 355
QY 266 EMFDKLYLSQGCIFKGVVYTNLIPKGLGLHCPTYNHPADPIIEVASGEYGDNL-PML 324
Db 356 HRFDKLVLSRGLLYFGKASEAMSYFSSIGCSPLAMPAPAEFLDLVNGNNDISVPSA 415
QY 325 FRA- - - - -VQNGLC- - - - -AMAEKKSPKEVNPAPCP- - - - -PCPP 356
Db 416 LKEKMKIIRLELYVRNVKCDVETQYLEEAYKQIAVMKMKLMAVPPLDEVKLMTCP- 474
QY 357 EVDPIESHFANSTLTQFCILKPKRTFSLRTVTLHRLPMSSHVTVGLGLYLHIGDD 416
Db 475 - - - - -KRWGSLSWMEQYCLLSLRG- - - - -IKERRHDYFSLWRVTVLSTAIILGLWQ-SDI 526
QY 417 ASKVFNNTGCLFSLMLFMAALMPTVLTFFLEMAVEMREHLNWTSLKAYLAKTMADV 476
Db 527 TSQRPRTSGLLFIYVWGFPPVFTAITFPQERAMLSKERESNMVLSAYFVARTSDI 586
QY 477 PFQVCPVVCYSIVYWMTGOPAEFSRFLFSALATATATAVAQSLGLLGAASNSLOVATF 536

RESULT 9

T21109
hypothetical protein F19B6.4 - Caenorhabditis elegans
C: Species: Caenorhabditis elegans
C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C: Accession: T21109
R: Thomas, K.
submitted to the EMBL Data Library, February 1996
A: Reference number: Z19375
A: Status: preliminary; translated from GB/EMBL/DBDJ
A: Molecule type: DNA
A: Residues: 1-695 <WIL>
A: Cross-references: EMBL:Z69635; NID:g1200023; PIDN:CAA93461.1; GSPDB:GN00022; CESP:F
A: Experimental source: clone F19B6
C: Genetics:
A: Gene: CESP:F19B6.4
A: Map position: 4
A: Introns: 9/1; 51/3; 167/2; 257/3; 515/3; 561/1
C: Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 20.5%; Score 686; DB 2; Length 695;

Best Local Similarity 27.5%; Pred. No. 1.4e-45;

Matches 184; Conservative 130; Mismatches 279; Indels 76; Gaps 17;

QY 31 PVL- - - - -TTHLKKVENHITAEQRFSLPKRSV- - - - -DIEFVLSYSVR- - - - - 71
Db 45 PVLINDTPVASIEQVEKELAHSSERFRPKPLSVTNDTESARIPSOVKTNAKKLVQNI 104
QY 72 EGPCWRKRGYK-TLKLCLSGKFCRRELIGIMGPGSAGKSTFMNLAGYRESGKM--GQIL 128
Db 105 QAVLVKKGVROELKIDGVARPGELTFIMGSSGAGKTTLLNLTGRNLKNIETGDIM 164
QY 129 VNGRPELRTFRKMSCYIMODMLPHLTVLEAMVYSANLNTENPDV- - - - -KNDLVTE 183
Db 165 INGRNMISNEMKKLSAYVOODDVFICTLTVRTLRFPAKL- - - - -RSPSALGATELDSIVDE 221
QY 184 ILTALGLMSCSHTRTAL- - - - -LSGQQRKRLAIALLELVNPPVPMFDEPTSGLDASCFQ 238
Db 222 LLVMSLKKCENTKVGTMTEKSLRGERKRLAFACELITDPPILFCDEPTSGLDSPMSHQ 281
QY 239 VYSLMKSLAOGGRTICTTHQPSAKLFEMFDKLYLSQGCIFKGVVYTNLIPKGLGLGH 298
Db 282 VIKALRQLTIEGTVICTTHQPSVSYVHMADQLLSQGHVAYAGPAKVDAFFRCGVP 341
QY 299 CPTYHNPADPIIEVASGE- - - - -YGDNLNPLMFRAVQNGLCMAEKKS- - - - -SPKNEVPAP 350
Db 342 IPKFVSSPDHFRMVRISHKSFETEDDYNKRIEKTIVLEHDMKKEQSTHSTLSSRRRHET 401
QY 351 CPPCPPEVDPISHTEATLTQFCILKPKRTFSLRTVTLHRLPMSSHVTVGLGLYL 410
Db 402 AP- - - - -FTFPRTWTAQFFIFORSSQLWRSRVLLVKLIQTLIIMSILCSTY 450
QY 411 LHIGDASKVFNNTGCLFSLMLFMAALMPTVLTFFLEMAVEMREHLNWTSLKAYILA 470
Db 451 YGLEIDKSLPSFKGFAFVSQVMHMLFMNPAMTVFWDYPPVVVREFQANMYSPSAYILA 510
QY 471 KTMADVPFQVCPVVCYSIVYWMTGOPAE- - - - -SRFLFSALATATATAVAQSLGLLIGA 526
Db 511 KTTADSIQVLPVFPVIFSGILLGNTSLPYSVVIITNYLIINLLSNACSVGSGFAAMCGH 570

RESULT 14

T34391

hypothetical protein T26A5.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000

C:Accession: T34391

R:Du, Z.

submitted to the EMBL Data Library, April 1994

A:Description: The sequence of C. elegans cosmid T26A5.

A:Reference number: Z21516

A:Accession: T34391

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-608 <DUZ>

A:Cross-references: EMBL:U00043; PIDN:AAC77504.1; GSPDB:GN00021; CESP:T26A5.1

A:Introns: 23/1; 96/3; 243/1; 342/2; 374/3; 403/1; 428/2; 464/3; 494/3; 534/2

A:Experimental source: strain Bristol N2; clone T26A5

C:Genetics:

A:Gene: CESP:T26A5.1

A:Map position: 3

A:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 19.0%; Score 635.5; DB 2; Length 608;

Best Local Similarity 27.1%; Pred. No. 1e-41;

Matches 156; Conservative 126; Mismatches 273; Indels 21; Gaps 6;

QY 82 KTLKCLSGKFCRRELIGIMPGSGAGKSTFMILAGYRESGM--KGQILVNGRPRELRTF 139

DB 43 RLLLNVSQYAKSGELLALMGAGAGKTYLLNMLCRNLKLGSTEGTIVNGN-EMAHKI 101

QY 140 RMKSYIMODMLLPHLTVLEAMVYSANLNTENPDVKNLVTALGLMSCSHTRTA 199

DB 102 SSISGFAQOEELFVGLTVKYLQIAKLRINGSKKLRDRVTVLHQLKWKCRDSKIG 161

QY 200 L-----LSGGQRKLALALELVNPPVFMFFDEPTSGLDSCFCFVQVSLMKSLAOGRTI 253

DB 162 VIGKKGISGGARLRTFACEMSLNPGLLFADEPTGLDSEMAESVIQILKGIKTRTI 221

QY 254 ICTIHQPSAKLFEMFDKLYLSOGQCIFKGVVNTLIPYLKGLGHCPTVHPADFTI--- 310

DB 222 ICTIHQPSQLYQMFHRYIYLANGSTAFQGTQESISFEKCHRVDPDEYNSEWLIYKL 281

QY 311 EVASGEYDNLPMLEFVAVQVGLCAMAEEKSPKEVPAFCPPPEVDPIESHTFATST 370

DB 282 AVQPGQKQSDRIQKIVEQEDSDHQVRMEQLSDVSEKIP--PPEM-----HRANV 332

QY 371 LQFCILPKRTLSILRTVTLHFRFMSHVIGVIGLGLYHIGDDASKVFNTCLFFS 430

DB 333 FTQIFALSTRCGIDVWRAPQLKAVIQKILFGLFGLLYLRTPYDARGIHNGALFFL 392

QY 431 MLFLMFAALMPTVLTFFLEMAVFMREHLNLYSLKAYYLAKTMADVPPFVQVCPVYCSIV 490

DB 393 AGEYIYSTAYAIMFFELNNEFFLVAREYHDLNLYTYFARCLSLIPFSTGILLIFV 452

QY 491 YNMTQPAETSFELFSLATATATVAOSGLLIGAANSLSQVATFVGVPTAIPVLLFSG 550

DB 453 YWLGILNTSMQVIVASITVILASQAASAFGIAMSCIFPTAQMATAVMAASPVLVLERFEG 512

QY 551 FVFSFKTPTTYLOWSSYLSVRYGEGVLTITTYGMERDGLTCLEERCPPREPQSTILRALD 610

DB 513 LYGNNTTTPAARLWLOWISWYFAFEGVLVNWQSEIDDFHSNAKNWNTSTINDVLDYFA 572

QY 611 VEDAKLYMDFLVIGLIFFLALRLAYLVLYRYVKSR 646

DB 573 FSESAPLDIIGLILISLAFYLGIFALVHRMKAR 608

RESULT 15

T04229

ABC-type transport protein F14M19.30 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 17-Mar-2000

C:Accession: T04229

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Hoheisel, J.; submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15262

A:Accession: T04229

A:Molecule type: DNA

A:Residues: 1-577 <BEV>

A:Cross-references: EMBL:AL049480

A:Experimental source: cultivar Columbia; BAC clone F14M19

C:Genetics:

A:Map position: 4

A:Note: F14M19.30

C:Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 18.7%; Score 626; DB 2; Length 577;

Best Local Similarity 29.1%; Pred. No. 5.4e-41;

Matches 166; Conservative 114; Mismatches 235; Indels 56; Gaps 13;

QY 96 ELIGMPSGAGKSTFMILAGYRESGMKGQILVNGRPRELRTFRKMSCYIMQDDMLLPH 155

DB 42 QILAIIGPSGAGKSTLLDILAA-RTSPSGSILLNSVLINPSSYKRKISSYVPQHDFTFPL 100

QY 156 LTVLEAMVYSANLNTENPDVKNLVTALGLMSCSHTRTAL--LSGGQRKLALIALE 214

DB 101 LTVSETFTFSASLLPKNLKSVSVASLLKELNTHLAHRLGQGLSGGERRRVSIGLS 160

QY 215 LVNPNPVFMFFDEPTSGLDSCFCFVQVSLMKSLAOG-GRITICTIHQPSAKLFEMFDKLYI 273

DB 161 LHDDEVLLDDEPTSGLDSCFSADFVQILKSTATSRERIVILSIHQPSFKILSLIDRLVL 220

QY 274 LSOGQCIFKGVVNTLIPYLKGLGHCPTVHPADFTIEVASGEYDNLPMLEFVAVQVGLC 333

DB 221 LSKGTIVHGRDLLEAFLLSKGFTVPOLNSLEVAMEI-----LQN--- 262

QY 334 AMAEKKSSPEKN---EVPAPCPPEVDPIESHTFATSTLQFCILPKRTLSILRTV 390

DB 263 -----IRDPYENANIALPDHCPESKKQKQKQSVIRYKSRITETISLSRREWKIIVTRQ 317

QY 391 LTHLRFMSHVIGVIGLGLYHIGDDASKVFNTCLFFSMLFMAALMPTVLTFFL--- 448

DB 318 LLLTNILSLVGLVGLVGLTYLNLGKRGKIRKRGFLFAFTLFL-----LSSTQTLPIFI 373

QY 449 -EMAVFMREHLNLYSLKAYYLAKTMADVPPFVQVCPVYCSIVYVMTQPAETSRFLFS 507

DB 374 DERPILLRETSSGLYRLSSHLANTLVFLPYLLLIATIIYSVSLVGLVGLCFSMQALAVFV 433

QY 508 ALATATVAOSGLLIGAANSLSQVATFVGVPTAIPVLLFSGFFVSPKTIPTTYLOWSSY 567

DB 434 LIVITIVLMANSFVLSLAPNYTAGTSSVTILLAAFFLFSGYFISKESLPKYLWLFYF 493

QY 568 LSVYRYGEGVLTITTYGMERDGLTCLEERC-PPREPQSTILRAL---DV-----EDAK 615

DB 494 FSWYKVALDALLINEY-----SCLHNKCLVWFEEASVNSCLVTGGVDLKNGLHQROR 546

QY 616 LYMDFLVIGITFLAURLLAYLVLYRYVKSR 646

DB 547 WFNVMILG-FFVLYRVLCFLVLLKRVSGSK 576

Search completed: June 10, 2003, 16:47:19

Job time : 47 secs

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	3347	100.0	646	9	US-10-072-621-9
2	3347	100.0	646	9	US-10-090-455-2
3	3301	98.6	646	9	US-10-154-452-4
4	3293	98.4	646	9	US-10-154-452-8
5	3287	98.2	646	9	US-10-090-455-13
6	3213	96.0	627	9	US-10-090-455-8
7	2356	70.4	674	9	US-10-090-455-4
8	2350	70.2	638	9	US-10-072-621-10
9	879.5	26.3	663	9	US-10-108-605-245
10	719	21.5	657	10	US-09-866-866A-14
11	717	21.4	655	10	US-09-866-866A-27
12	712	21.3	655	9	US-09-981-353-35
13	712	21.3	655	9	US-10-120-687-61
14	710	21.2	655	9	US-09-961-086-1
15	704	21.0	655	9	US-10-090-455-5
16	704	21.0	655	10	US-09-866-866A-10
17	624.5	18.7	604	10	US-09-745-763-197
18	585.5	17.5	651	9	US-10-090-455-6
19	585.5	17.5	651	9	US-09-989-981A-6

```

US-10-072-621-9
; Sequence 9, Application US/10072621
; Patent No. US20020169137A1
; GENERAL INFORMATION:
; APPLICANT: Reiner, Peter B.
; APPLICANT: Connop, Bruce P.
; APPLICANT: Pollard, Michelle
; TITLE OF INVENTION: REGULATION OF
; FILE REFERENCE: 100103.402
; CURRENT APPLICATION NUMBER: US/10/
; CURRENT FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Vers
; SEQ ID NO 9
; LENGTH: 646
; TYPE: prt
; ORGANISM: Homo sapiens
US-10-072-621-9

```

Query Match	100.0%;	Score 3347;	DB 9;	Length 646;
Best Local Similarity	100.0%;	Pred. NO. 1.3e-293;		
Matches 646;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAEKALEAVGCGIGPGAVAMAVTLEDCAEPVLTTHLKVKVENHITTEAQRFSHLPKRAVD	60	
DB	1	MAEKALEAVGCGIGPGAVAMAVTLEDCAEPVLTTHLKVKVENHITTEAQRFSHLPKRAVD	60	
QY	61	IEFVELSYSVREGPCWRKRGYKTYLLKCLSKGFCRRRELIGMPGSGACKSTFMNLAGYRE	120	
DB	61	IEFVELSYSVREGPCWRKRGYKTYLLKCLSKGFCRRRELIGMPGSGACKSTFMNLAGYRE	120	
QY	121	SGMKGOILVNGRPRELRTFRKMSCYIMODMLLPHLTVLEAMMYSANLITENPDVNKDL	180	
DB	121	SGMKGOILVNGRPRELRTFRKMSCYIMODMLLPHLTVLEAMMYSANLITENPDVNKDL	180	
QY	181	VTEILTALGLMSCSHRTALLSSGQQRRLAIAELVNNPVPVFDEPTSGLDASCFQVV	240	
DB	181	VTEILTALGLMSCSHRTALLSSGQQRRLAIAELVNNPVPVFDEPTSGLDASCFQVV	240	
QY	241	SLMKSLAOGERTICTIHOPSAKLFEFMFDKLYILSOGOCIFKGVVTNLIPVLCGLGLHCP	300	

Db 241 SLKSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVNLIPYLKGLGHCP 300
Qy 301 TYHNPADFIIEVASGEYGLNPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
Db 301 TYHNPADFIIEVASGEYGLNPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
Qy 361 IESHTATSTLTQFCILFKRTFSLILRDVTLTHLRMSHVIVIGVLGLLHLHIGDDASKV 420
Db 361 IESHTATSTLTQFCILFKRTFSLILRDVTLTHLRMSHVIVIGVLGLLHLHIGDDASKV 420
Qy 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Db 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Qy 481 VCPVYCSIVYWMGTGPAETSRELLFSALATATVAQSLGLLGAASNSLQVATFVGPV 540
Db 481 VCPVYCSIVYWMGTGPAETSRELLFSALATATVAQSLGLLGAASNSLQVATFVGPV 540
Qy 541 TAIPVLLFSGFFVSFKTPTIYLOWSSYLSVRYGFEGLVLTTCLEERCPCR 600
Db 541 TAIPVLLFSGFFVSFKTPTIYLOWSSYLSVRYGFEGLVLTTCLEERCPCR 600
Qy 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLYRVKSR 646
Db 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLYRVKSR 646

RESULT 2

US-10-090-455-2

; Sequence 2, Application US/10090455

; Publication No. US20030027259A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Hongyun

; APPLICANT: Le Bihan, Stephane

; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF

; FILE REFERENCE: 100103.406

; CURRENT APPLICATION NUMBER: US/10/090,455

; CURRENT FILING DATE: 2002-03-01

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 646

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-090-455-2

Query Match 100.0%; Score 3347; DB 9; Length 646;
Best Local Similarity 100.0%; Pred. No. 1.3e-293;
Matches 646; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSAYD 60
1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSAYD 60
61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRE 120
61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRE 120
121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSGLDSASCFQV 240
181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSGLDSASCFQV 240
241 SLKMSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVNLIPYLKGLGHCP 300
241 SLKMSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVNLIPYLKGLGHCP 300
301 TYHNPADFIIEVASGEYGLNPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360

Db 301 TYHNPADFIIEVASGEYGLNPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
Qy 361 IESHTATSTLTQFCILFKRTFSLILRDVTLTHLRMSHVIVIGVLGLLHLHIGDDASKV 420
Db 361 IESHTATSTLTQFCILFKRTFSLILRDVTLTHLRMSHVIVIGVLGLLHLHIGDDASKV 420
Qy 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Db 421 FNNTGCLFFSMLFMAALMPTVLTTPLEMAVFMREHLNLYWYSLKAYYLAKTMADVPFQV 480
Qy 481 VCPVYCSIVYWMGTGPAETSRELLFSALATATVAQSLGLLGAASNSLQVATFVGPV 540
Db 481 VCPVYCSIVYWMGTGPAETSRELLFSALATATVAQSLGLLGAASNSLQVATFVGPV 540
Qy 541 TAIPVLLFSGFFVSFKTPTIYLOWSSYLSVRYGFEGLVLTTCLEERCPCR 600
Db 541 TAIPVLLFSGFFVSFKTPTIYLOWSSYLSVRYGFEGLVLTTCLEERCPCR 600
Qy 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLYRVKSR 646
Db 601 EPOSILRALDVEDAKLYMDFLVLGIFFLALRLAYLVLYRVKSR 646

RESULT 3

US-10-154-452-4

; Sequence 4, Application US/10154452

; Publication No. US20020192821A1

; GENERAL INFORMATION:

; APPLICANT: Reiner, Peter B.

; APPLICANT: Roy, Josee

; APPLICANT: Connop, Bruce P.

; TITLE OF INVENTION: INCREASED FUNCTIONAL ACTIVITY AND/OR

; TITLE OF INVENTION: EXPRESSION OF ABC TRANSPORTERS PROTECTS AGAINST THE LOSS OF

; TITLE OF INVENTION: DOPAMINE NEURONS ASSOCIATED WITH PARKINSON'S DISEASE

; FILE REFERENCE: 100103.420

; CURRENT APPLICATION NUMBER: US/10/154,452

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 9

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 646

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-154-452-4

Query Match 98.6%; Score 3301; DB 9; Length 646;
Best Local Similarity 98.8%; Pred. No. 1.9e-289;
Matches 638; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSAYD 60
1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKKVENHITAEQRFSLPKRSAYD 60
61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRE 120
61 IEVELSYSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRE 120
121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
121 SGMKGOILVNGRPRELRTFRKMSCYIMQDDMLPLHLTVLEAMVMSANLNTENPDVKNL 180
181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSGLDSASCFQV 240
181 VTEILTALGLMSCSHRTALLSGQQRKRLAIALELVNPNPVMFEDEPTSGLDSASCFQV 240
241 SLKMSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVNLIPYLKGLGHCP 300
241 SLKMSLAQGGRTICTIHQPSAKLFEMFDKLYILSOGQCFKGVVNLIPYLKGLGHCP 300
301 TYHNPADFIIEVASGEYGLNPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360
301 TYHNPADFIIEVASGEYGLNPMFLRAVQNGLCAMAEKSSPEKNEVPAPCPPEVDP 360

```
QY 361 IESHTFATSTLTQRCILFKRTFLSILRDTVLTTHLRFMSHVIVGLIGLLYHIGDDASKV 420
Db 361 IESHTFATSTLTQRCILFKRTFLSILRDTVLTTHLRFMSHVIVGLIGLLYHIGDDASKV 420
QY 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
Db 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
QY 481 VCPVYCSIVYVMTGQPAETSRLFFSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
Db 481 VCPVYCSIVYVMTGQPAETSRLFFSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
QY 541 TAIPIVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
Db 541 TAIPIVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
QY 601 EPOQILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSER 646
Db 601 EPOQILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSER 646

RESULT 4
US-10-154-452-8
; Sequence 8, Application US/10154452
; Publication No. US20020192821A1
; GENERAL INFORMATION:
; APPLICANT: Reiner, Peter B.
; APPLICANT: Roy, Josee
; APPLICANT: Connop, Bruce P.
; TITLE OF INVENTION: INCREASED FUNCTIONAL ACTIVITY AND/OR
; TITLE OF INVENTION: EXPRESSION OF ABC TRANSPORTERS PROTECTS AGAINST THE LOSS OF
; TITLE OF INVENTION: DOPAMINE NEURONS ASSOCIATED WITH PARKINSON'S DISEASE
; FILE REFERENCE: 100103.420
; CURRENT APPLICATION NUMBER: US/10/154.452
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-154-452-8

Query Match 98.4%; Score 3293; DB 9; Length 646;
Best Local Similarity 98.5%; Pred. No. 9.9e-289;
Matches 636; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

1 MAEKALAEVCGGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
1 MAEKALAEVCGGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLPHTLVLEAMVYSANLNTENPDVKNL 180
121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLPHTLVLEAMVYSANLKLSEKQVKEL 180
181 VTEILTALGLMSCSHTRTALLSGGQRKRLAIALELVNPNPVMFFDEPTSGLDASCFQV 240
181 VTEILTALGLMSCSRTTALLSGGQRKRLAIALELVNPNPVMFFDEPTSGLDASCFQV 240
241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
301 TYHNPADFIEVASGEYGDNLNPLFRAVQNGLCMAEKKSSPEKNEVPAPCPPEVDP 360
301 TYHNPADFIEVASGEYGDNLNPLFRAVQNGLCMAEKKSSPEKNEVPAPCPPEVDP 360
361 IESHTFATSTLTQRCILFKRTFLSILRDTVLTTHLRFMSHVIVGLIGLLYHIGDDASKV 420
Db 361 IESHTFATSTLTQRCILFKRTFLSILRDTVLTTHLRFMSHVIVGLIGLLYHIGDDASKV 420
```

```
Db 361 IESHTFATSTLTQRCILFKRTFLSILRDTVLTTHLRFMSHVIVGLIGLLYHIGDDASKV 420
QY 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
Db 421 FNNTGCLFFSMLFMAALMPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPOV 480
QY 481 VCPVYCSIVYVMTGQPAETSRLFFSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
Db 481 VCPVYCSIVYVMTGQPAETSRLFFSALATATATVAQSLGLLGAASNSLQVATFVGPV 540
QY 541 TAIPIVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
Db 541 TAIPIVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTYGMERGDLTCLERCPCR 600
QY 601 EPOQILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSER 646
Db 601 EPOQILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSER 646

RESULT 5
US-10-090-455-13
; Sequence 13, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: 579, 598
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-090-455-13

Query Match 98.2%; Score 3287; DB 9; Length 646;
Best Local Similarity 98.5%; Pred. No. 3.5e-288;
Matches 636; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

1 MAEKALAEVCGGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
1 MAEKALAEVCGGLGPGAVAMAVTLEDGAEPVLTTHLKKVENHTEAQRFSHLPKRSAVD 60
61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
61 IEFVELSVSVREGPCWRKRGYKTLKCLSGKFCRRELIGIMPGSGAGKSTFMNLAGVRE 120
121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLPHTLVLEAMVYSANLNTENPDVKNL 180
121 SGMKQILVNGRPRELRTFRKMSCYIMQDDMLPHTLVLEAMVYSANLKLSEKQVKEL 180
181 VTEILTALGLMSCSHTRTALLSGGQRKRLAIALELVNPNPVMFFDEPTSGLDASCFQV 240
181 VTEILTALGLMSCSRTTALLSGGQRKRLAIALELVNPNPVMFFDEPTSGLDASCFQV 240
241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
241 SLKSLAOGGRTIICTHOPSAKLFEMFDKLYILSOGOCIFKGVVVTNLIPLYKGLGLHCP 300
301 TYHNPADFIEVASGEYGDNLNPLFRAVQNGLCMAEKKSSPEKNEVPAPCPPEVDP 360
301 TYHNPADFIEVASGEYGDNLNPLFRAVQNGLCMAEKKSSPEKNEVPAPCPPEVDP 360
361 IESHTFATSTLTQRCILFKRTFLSILRDTVLTTHLRFMSHVIVGLIGLLYHIGDDASKV 420
Db 361 IESHTFATSTLTQRCILFKRTFLSILRDTVLTTHLRFMSHVIVGLIGLLYHIGDDASKV 420
```

Qy 421 FNNTCGLFFSMLFLMFAALMPTVLTFFPLEMAVFMREHLNWNYSKAYYLAKTMADVPFQV 480
Db 421 FNNTCGLFFSMLFLMFAALMPTVLTFFPLEMAVFMREHLNWNYSKAYYLAKTMADVPFQV 480
Qy 481 VCPVVCYSIVYVMTGQPAETSRFLFSALATATATVAQSLGLLIGAASNSLQVATFYGPV 540
Db 481 VCPVVCYSIVYVMTGQPAETSRFLFSALATATATVAQSLGLLIGAASNSLQVATFYGPV 540
Qy 541 TAIPVLLFSGFFVSFKTIPYIYLQWSSYLSYVRGFEQVILTYIYMERGDLTCLERCPFR 600
Db 541 TAIPVLLFSGFFVSFKTIPYIYLQWSSYLSYVRGFEQVILTYIYMERGDLTCLERCPFR 600
Qy 601 EPOSILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSR 646
Db 601 EPOSILRALDVEDAKLYMDLVLGIFFLALRLAYLVLRYVKSR 646

RESULT 6

US-10-090-455-8
; Sequence 8, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-8

Query Match 96.08; Score 3213; DB 9; Length 627;
Best Local Similarity 98.98; Pred. No. 1.6e-281;
Matches 620; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Qy 20 MAVTLEDGAPPVLTTHLKKVENHITAEQRFSLPKRSADVIEFVELSYSVREGPCWKRKR 79
Db 1 MAVTLEDGAPPVLTTHLKKVENHITAEQRFSLPKRSADVIEFVELSYSVREGPCWKRKR 60
Qy 80 GYKTLKCLSKFCRRELIGIMPGSGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTF 139
Db 61 GYKTLKCLSKFCRRELIGIMPGSGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTF 120
Qy 140 RKMSCYIMQDDMLLPHLTVLEAMMVSANLNTENPDVKNLVTILTALGLMSCSHTRTA 199
Db 121 RKMSCYIMQDDMLLPHLTVLEAMMVSANLNTENPDVKNLVTILTALGLMSCSHTRTA 180
Qy 200 LLSGGQKRRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHQ 259
Db 181 LLSGGQKRRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHQ 240
Qy 260 PSAKLFEWFKLYTLISOGQICIFKGVVNTLIPYLKGLGHCPTVHNPADEFIEVASGEYGD 319
Db 241 PSAKLFEWFKLYTLISOGQICIFKGVVNTLIPYLKGLGHCPTVHNPADEFIEVASGEYGD 300
Qy 320 LNPMLFRAVONGLCAMAEKSSPEKNEVPAPCPPEVDPPIESHTATSTLTQFCILFK 379
Db 301 LNPMLFRAVONGLCAMAEKSSPEKNEVPAPCPPEVDPPIESHTATSTLTQFCILFK 360
Qy 380 RTFLSLRDVTLHLRFMSHWVIGVLGLYHLIGDDASQVFNNTGCLFTSMFLMFAAL 439
Db 361 RTFLSLRDVTLHLRFMSHWVIGVLGLYHLIGDDASQVFNNTGCLFTSMFLMFAAL 420
Qy 440 MPTVLTFLPMAVFMREHLNWNYSKAYYLAKTMADVPFQVCPVVCYSIVYVMTGQPAE 499
Db 421 MPTVLTFLPMAVFMREHLNWNYSKAYYLAKTMADVPFQVCPVVCYSIVYVMTGQPAE 480

Qy 500 TSRELLFSALATATVAQSLGLLIGAASNSLQVATFYGPVTAIPVLLFSGFFVSFKTIP 559
Db 481 TSRELLFSALATATVAQSLGLLIGAASNSLQVATFYGPVTAIPVLLFSGFFVSFKTIP 540
Qy 560 TYLOWSSYLSVRYVGFEGVILTYIYMERGDLTCLERCPREPOSILRALDVEDAKLYMD 619
Db 541 TYLOWSSYLSVRYVGFEGVILTYIYMERGDLTCLERCPREPOSILRALDVEDAKLYMD 600
Qy 620 FLVLGIFFLALRLAYLVLRYVKSR 646
Db 601 FLVLGIFFLALRLAYLVLRYVKSR 627

RESULT 7

US-10-090-455-4
; Sequence 4, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-4

Query Match 70.48; Score 2356; DB 9; Length 674;
Best Local Similarity 70.18; Pred. No. 4.6e-204;
Matches 454; Conservative 80; Mismatches 76; Indels 38; Gaps 4;
Qy 26 DGAEPVLTTHLKKVENHITAEQRFSLPKRSADVIEFVELSYSVREGPCWKRKYKTL 85
Db 38 EATETDLNGLHKKVDNLTAEQRFSLPRAAVNIEFRLDSYVPEGPMWRKKGKYL 97
Qy 86 KCLSGKFCRRELIGIMPGSGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTFRKMSY 145
Db 98 KGISGKFNSELVAIMGPGSGAGKSTFMNLAGYRETGMKGAVALINGLPRDLRCFRKVS 157
Qy 146 IMQDDMLLPHLTVLEAMMVSANLNTENPDVKNLVTILTALGLMSCSHTRTALLSGQ 205
Db 158 IMQDDMLLPHLTVOEAMMVSANLNTENPDVKNLVTILTALGLMSCSHTRTALLSGQ 217
Qy 206 RKRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHQPSAKLF 265
Db 218 RKRLAIALELVNPNVPMFFDEPTSGDLSASCFCQVVSMLKSLAQGGRTIICTHQPSAKLF 277
Qy 266 EMFDKLYLSGGQICIFKGVVNTLIPYLKGLGHCPTVHNPADEFIEVASGEYGDNLPMFL 325
Db 278 ELFDOLYLSGGQICIFKGVVNTLIPYLKGLGHCPTVHNPADEFIEVASGEYGDNLPMFL 337
Qy 326 RAVONGLCAMAEKSSPEKNEVPAPCPPEVDPPIESHTATSTLTQFCILFK 361
Db 338 RAVREGMC-----DSHKKDLGGDA-----EVNPFVLRHPSSEVVKOTKRLKURKDKSSS 386
Qy 362 --ESHTATSTLTQFCILFKRTFLSLRDVTLHLRFMSHWVIGVLGLYHLIGDDASK 419
Db 387 MEGCHSFSASCLTQFCILFKRTFLSLRDVTLHLRFMSHWVIGVLGLYHLIGDDASK 446
Qy 420 VFNNTGCLFTSMFLMFAALMPTVLTFLPMAVFMREHLNWNYSKAYYLAKTMADVPFQ 479
Db 447 VLSNSGFLFFSMLFLMFAALMPTVLTFLPMAVFMREHLNWNYSKAYYLAKTMADVPFQ 506
Qy 480 VPCPVVCYSIVYVMTGQPAETSRFLFSALATATVAQSLGLLIGAASNSLQVATFYGP 539
Db 507 IMPVVCYSIVYVMTGQPAETSRFLFSALATATVAQSLGLLIGAASNSLQVATFYGP 566
Qy 540 VTAIPVLLFSGFFVSFKTIPYIYLQWSSYLSYVRGFEQVILTYIYMERGDLTCLERCP 598

Db 595 INOWADVEGEISCTSSNTTCP--SSGKVIETLNFSAADPLDYVGLAILVSRVLAYL 653
QY 637 VLRYRVK 643
Db 654 AULRLAR 660

RESULT 10
US-09-866-866A-14
; Sequence 14, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 14
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-866-866A-14

Query Match 21.5%; Score 719; DB 10; Length 657;
Best Local Similarity 29.1%; Pred. No. 3.3e-56;
Matches 183; Conservative 133; Mismatches 258; Indels 54; Gaps 16;
QY 61 IEFVELSY--SVREGPCWRKRGYKTLKCLSGKFCRRELIGMGPSGAGKSTFMNLAGY 118
Db 37 LSFHHITYVKVKGSLVRKTEKEILSDING-IMKPGLNAILGPTGGKGSLLDVLAR 95
QY 119 RE-SGMKGQILVNGRPRELRTFRKMSCYIMQDDMLPLHPLTVLEAMVVSANLN--TENPD 175
Db 96 KDPKGLSGDVLNGAPQPAH-FKCCSGYVQDDVVMGTLTVRENLOFSAALRLPTMKNH 154
QY 176 VKNDLVTETLTALGMLSCSHTRTAL-----LSGQGRKRLAIALELVNPNPVMFDEPTSG 230
Db 155 ENERINTIIEKLGLEKVDKVGTOFIRGIGSGGERKRTSIGMELTIDPSILFLDEPTG 214
QY 231 LDSASCFQVYSLMKSLAOGGRTIICTIHOPSAKLFEMEDKLYILSQGQCIFKGVVNTLIP 290
Db 215 LDSSTANAVLLLLKRMKSKOGRTIIFSIHQPRYSIFKLFDSLTLLASGKLVFHGPAQKALE 274
QY 291 YLKGGLGHCPTVHNPADEFIEVASGEYGDNLPMFLFRAVQNGLCAMAEKKSPEKNEVP-- 348
Db 275 YFASAGYHCEPYNNPADFLLVINGDSSAV--MLAREEQDN---EANKTEEPSKGEKPI 329
QY 349 -----APCPPCPPPEVDPI-----ESHFTATSTLTQFCILEKRTFLSTL 386
Db 330 ENLSFEYINSALYGETKAELODQPGAQEKKGTSAPKEPVYVTSFCHQLRWIARRSFKNLL 389
QY 387 RDTVLTLHFRMSHVIGVLIGLGLYHLIGDDASKVFNNTGCLFFSMLFLMFAALMPTVLF 446
Db 390 GNPOASVAQLIIVTILGLIIGAIYFDLKYDAAGMONRAGVLFLITNOCFSSVASVEL-F 448
QY 447 PLEMAVFMREHLNRYVSLKAYLAKTMAD-VPFQVVCVVYCSIVVWMTGQPAETSRLFL 505
Db 449 VVEKFLFIHEYISGYRVSSYFFGKVMSDLPMRELPVSIFTCILYFMLGLKKTVDAPFI 508
QY 506 FSALATATVAOSGLGLLIGAASNSLOVATFVPVTAIPVLLFSGFFVSKTIPYLOW 565
Db 509 MMFTLIMWAYTASSMALATATQSVSVATLMTWTAFAVFMFLFSGLLVNERTIGPWSL 568
QY 566 SYLSYVRVGFEGVILTIY-GME-----RGDLTCLER--CPFREPOQILRALDVEDAK 615

Db 569 QYFSIPRYGFTALQYNEFLGQEPFCGFNVNTDSTCVNSYAICTGNE-YLINQOIELSPWG 627
QY 616 LYMDFLVLGIFFLALRLALLAYLVLRVYK 643
Db 628 LMKNHVALACMIIFLTIAYLKLLFLKK 655

RESULT 11
US-09-866-866A-27
; Sequence 27, Application US/09866866A
; Patent No. US20020102244A1
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 27
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-27

Query Match 21.4%; Score 717; DB 10; Length 655;
Best Local Similarity 28.0%; Pred. No. 4.9e-56;
Matches 178; Conservative 141; Mismatches 245; Indels 72; Gaps 16;
QY 61 IEFVELSYSV--REG--PCWRKRGYKTLKCLSGKFCRRELIGMGPSGAGKSTFMNILA 116
Db 37 LSFHHITYVKVKGSLVLPK-RKPVEKEILSNING-IMKPGLNAILGPTGGKGSLLDVL 94
QY 117 GYRE-SGMKGQILVNGRPRELRTFRKMSCYIMQDDMLPLHPLTVLEAMVVSANLN--TEN 173
Db 95 ARKDPGSLGCDVLNGAPRP-ANFKCNSGVVQDDVVMGTLTVRENLOFSAALRLATWT 153
QY 174 PDVKNDLVTETLTALGMLSCSHTRTAL-----LSGQGRKRLAIALELVNPNPVMFDEPT 228
Db 154 NHEKNERINRVIOELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELTIDPSILFLDEPT 213
QY 229 SGLDSASCFQVYSLMKSLAOGGRTIICTIHOPSAKLFEMEDKLYILSQGQCIFKGVVNTL 288
Db 214 TGLDSTANAVLLLLKRMKSKOGRTIIFSIHQPRYSIFKLFDSLTLLASGKLVFHGPAQ 273
QY 289 IYFLKGLGHCPTVHNPADEFIEVASGEYGDNLPMFLFRAVQNGLCAMAEKKSPEKNEVP 348
Db 274 LOYFESAGYHCEAYNNPADFLLIINGD-----STAVALNREEDFRATEII 319
QY 349 ACPPCPPPEVDPI-----ESHFTATSTLTQFCI 376
Db 320 EPKSKDKPLIEKLAETIYVNSFSYKETAELHQLSGEKKKKTIVFKETISYTTSFCHQLRW 379
QY 377 LFKRTFLSLTRDTVTLHFRMSHVIGVLIGLGLYHLIGDDASKVFNNTGCLFFSMLFLMF 436
Db 380 VSKRSFKNLLGNPQASIAQIIVTVVLGVLIGAIYFGLKNDSTGIONRAGVLFLITNOCF 439
QY 437 AALMPTVTLTFFLEMAVFMREHLNRYVSLKAYLAKTMAD-VPFQVVCVVYCSIVVWMTG 495
Db 440 SSVASVEL-FVVEKKLFIHEYISGYRVSSYFFGKVLKLLSLLPMRLPSIIFTCIVYFMLG 498
QY 496 QPAETSRLFLFSALATATVAOSGLGLLIGAASNSLOVATFVPVTAIPVLLFSGFFVSF 555
Db 499 LKAKADAFVFMFTLMWAYSSASSMALATAAGQSVSVATLMTWTAFAVFMFLFSGLLVN 558

556 KTTPTYLQWSSYLVYRGEGVLTIT-----GME-RDGLVCLERCBFRPQSIILR 607
QY
559 TTASLWLSLQYFSIPRYGTALQHNEFLGONFCPLGNATGNCNATAT- 617
Db
608 ALDVEDAKLYMDFVLVLGIFFLALRLLAYLVLYRVK 643
QY
618 GIDLSPWLKKNHVALACMIVIFITAIYKLLFLKK 653
Db

RESULT 12
US-09-981-353-35
; Sequence 35, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 35
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc:feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1
US-09-981-353-35

Query Match	21.3%	Score 712:	DB 9:	Length 655;
Best Local Similarity	27.8%	Pred. No. 1.4e-55;		
Matches 177:	Conservative 141;	Mismatches 246;	Indels 72;	Gaps 16;

QY	61	IEFVELSYSV--REG--PCWKRGYKTLKCLSGKCFRRELIGTMGPSGAGKSTFEMILA	116
DB	37	LSFHNICYRVKLSGFLPC-RKPVEKILSNG-IMKPGNALIGFTGGKSLDLVLA	94
QY	117	GYRE-SGMKQIILVNGRPRELTRFRKMSCYIMODMLLPHLTVLEAMVVSANLML--TEN	173
DB	95	ARKDPGLSGSDVLINGAPRP-ANFKNSGYVVDVYVNGTLTVRENLFQFSAALRLATMT	155
QY	174	PDVKNDLVTIELTALGLMCSHRTAL-----LSGGORKKRLAIALELVNPNPPVFFDEPT	228
DB	154	NHEKNERINAVIQELGLDKVADSKVGTOFTRGVSGGKKRTSIGMELITDPSILFDEPT	213
QY	229	SGLDSACFOVSLMKSLAGOGRTIICTIHOPSAKLFEMFDKLYLSOGOCIFKGVVYNL	288
DB	214	TGSDSTANAVLLLRMSKOGRTIIFSIHQPRYSIFKLFDSLTLLASGRIMFHGPAQEA	273
QY	289	IPYKLGLGLHCPYHNPADETFIEVASGEYGDLPNMLFRAVQNLCAAEKSSPEKNEVP	348
DB	274	LGYESAGYHCEAYNNPADEFLIDNGD-----STAVALNREEDFKATEII	319
QY	349	APCPCPPEVDPI-----ESH-----TFATSTLTQFCI	376
DB	320	EPSKQDKPLIEKLAIEYVNSFYKETAELHQLSGGEKKKITVFEIKSYTTSFCHQLRW	379
QY	377	LEKRTFLSLIURDVTVLTHLRFMSHWVIGVLGLLYLHIGDDASKVFNNTGCLFSFMLFMF	436
DB	380	VSRSEKNLLCNQASIAQIIVTVVLGEVIGAIYFGLKNDGSTGIONRAGVLFTTTNQCF	439
QY	437	AALMPTVLTFFLEMAVFMREHLNYSKAYLYAKTMAD--VPFOVGVVYVCISVYVMWG	495
DB	440	SSVSAREL-EVWEKKLPIHEYISGYRVSSVFLGKLLSLDLPMLPSIIITTCIVYFMLG	498
QY	496	QPAETSEFLFSALATATALVAOSLGLLIGNAANSLOVATFVGVPVTAIPVLVPSGFFVSP	555
DB	499	LKPKADAFFVMFTLMMAVVSASSMALAIAGOSVSVATLLMTICFVFMFIISGLLVNL	558
QY	556	KTIPTYLQWSYLSYRVYGFEGVITLTY-----GME-RGDLTCLCEERPQFRPQSILR	607

Db 559 TTIASWLSWLOYFSIPRYGTALQHNEFLQONFCPLNATGNNPCNYATCT-GEEYLVKQ 617
 Qy 608 ALDVEDAKLYMDFLVLGIFFFALRLRLAYLVLRYVK 643
 Db 618 GIDLSPWGLKNHVALACMIVIFLTITAYLKLFLKK 653

RESULT 13
 US-10-120-687-61
 : Sequence 61, Application US/10120687
 : Publication No. US20030082155A1
 : GENERAL INFORMATION:
 : APPLICANT: Massachusetts General Hospital
 : TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treati
 : TITLE OF INVENTION: Mellitus
 : FILE REFERENCE: 3284/1235B
 : CURRENT APPLICATION NUMBER: US/10/120,687
 : CURRENT FILING DATE: 2002-04-11
 : PRIOR APPLICATION NUMBER: US60/169082
 : PRIOR FILING DATE: 1999-12-06
 : PRIOR APPLICATION NUMBER: US 09/963,875
 : PRIOR FILING DATE: 2001-09-25
 : PRIOR APPLICATION NUMBER: US 60/215109
 : PRIOR FILING DATE: 2000-06-28
 : PRIOR APPLICATION NUMBER: US 60/238880
 : PRIOR FILING DATE: 2000-10-06
 : PRIOR APPLICATION NUMBER: US 09/731261
 : PRIOR FILING DATE: 2000-12-06
 : NUMBER OF SEQ ID NOS: 61
 : SOFTWARE: PatentIn version 3.1
 : SEQ ID NO 61
 : LENGTH: 655
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-10-120-687-61

Query Match	21.3%	Score 712	DB 9	Length 655
Best Local Similarity	27.8%	Pred. No. 1.4e-55		
Matches 177	Conservative 141	Mismatches 246	Indels 72	Gaps 16
QY	61	IEFVELSVSV--REG--PCWRKRGYKTLKCLSCGKFCRRELIGIMGSPGACKSTFMNLLA	116	
DB	37	LSFNICYRVKLKSGFLPC-RKPVEKEILSNING-IMKPGLNALIGTGGGKSLLDVLA	94	
QY	117	GYRE-SGMKGQILVNGRPRELRTFKNSCYITQDMILLPLHLTLEAMVNSANLNL--TEN	173	
DB	95	ARKDPGSLGSDVLINGARPP-ANFKCNSGYVQDDVVMGTLTVRENFQFSAALRLATMT	153	
QY	174	PDVKNDLVTEILTALGLMCSHTRTAL-----LSGGQKRKRLAIALELVNPNPVFFDEPT	228	
DB	154	NHEKNERINRVIQELGLDKVADSKVGTFIRGVSGGERKRTSIGMELITDPSILFDEPT	213	
QY	229	SGLDSASCFQVSLMKSLAQGRITIICTIHQPSAKLPEMFDPKLYLSQCQCIFKGVVTNL	288	
DB	214	TGLSDSTANAVALLLKRMSKOGRTIIFSIOHPRYSIFKLEPDSLTLASGRUMFHGAQEA	273	
QY	289	IPYLLKGLGLHCTPHYNPADFTIEVASGEGYDLPNMLFRAVQNGLCAMAEKSSPEKNEVP	348	
DB	274	LGYESAGYHCEAINNPADFELDIINGD-----STAVALNREEDFKATEII	319	
QY	349	APCPPCPEVDPI-----ESH-----TEATSLTQFCI	376	
DB	320	EPSKODKPLIEKLAELVYNSSFYKETAELHQLSGGEKKKIIVFKELSYTTSFCHQLRW	379	
QY	377	LFKRTFELSILRDTVLTHLRFNHSHVVGVLIGLLYHIGDDASKYFNNNTGCLFFSMLFMF	436	
DB	380	VKSRSFKNLGNPQASIAQIIVTVVLGVIGCAIVFGLKNDSTGIGNRAGVLFFLTNQCF	439	
QY	437	AALMPTVLTFFLENAVEMREHLNTWYSUKAYYLAKTHMAD-VPFQVQCVPYVCISVYWMTG	495	
DB	440	SSVSAVEL-FVVEKKLFTHIEYISGYRVSSYFLGKLLSDLLPMLRMLPSIIPTCTCIVFMLG	498	
QY	496	OPAESTRFLLESALATATALVAOSLGLLIGAANSLSLOVATEVGPVPTAIPVLVPSGFFVSF	555	

Db 499 LKPKADAFFVMMFTLMVAVSASSMALAIAAGOSVSVATLLMTICFVFMIFSGLLVNL 558
Qy 556 KTIPTYLOWSSYLSYRYGEGVILTYI-----GME-RGDLTCLCEERCPFRPQSILR 607
Db 559 TTIASWLSWLQYFESIPRYGFTALQHNFLQGNFCPGLNATGNNPCNYATCT-GEEYLKQ 617
Qy 608 ALQVEDAKLYMDFLVLGIFFLALRLAYLRLVRVK 643
Db 618 GIDLSPWGLKKNHVALACMIVIFLTIAYLKLLFLKK 653

RESULT 14,
US-09-961-086-1
; Sequence 1, Application US/09961086
; Publication No. US20030036645A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynne
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; FILE OF INVENTION: WHICH ENCODES IT
; FILE REFERENCE: EP19376-019
; CURRENT APPLICATION NUMBER: US/09/961.086
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1

Query Match 21.2%; Score 710; DB 9; Length 655;
Best Local Similarity 27.8%; Pred. No. 2.1e-55;
Matches 177; Conservative 140; Mismatches 247; Indels 72; Gaps 16;

Qy 61 IEVELSYV--REG--PCWRKRGYKTLKCLSGKFCRRELIGTMPSGAGKSTFMNLA 116
Db 37 LSFHNICRYVKLSGFLPC-RKPEKEILSNING-IMKPGNALIGPTGGKSSLLDVL 94

Qy 117 GYRE-SGMKQILVNGRPRELRTFRKMSCYIMODMLLPHLTVLEAMVMSANL--TEN 173
Db 95 ARKDPSGLSGDLVINGAPRP-ANFKCNSGVYVQDDVVMGTLTVRENLFQFSAALRLATTMT 153

Qy 174 PDVKNLVTLEILALGLMSCSHRTAL-----LSGGORKRLATALELVNPPVFFDEPT 228
Db 154 NHEKNERINRVIELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFDEPT 213

Qy 229 SGDSASCFVSVLSKSLAOGGRTIITIHQPSAKLFEMEDKLYLSOGOCIFKGVVNTL 288
Db 214 TGLDSSANAVLLLLKRMKSQGRITIFSIHQPRYSIFKLFDSLTLASGRMLFHGPAQEA 273

Qy 289 IPYKGLGLHCPVYHNPADFIIEVASGEYGDLPNMFRAVONGLCMAEKKSSPEKNEVP 348
Db 274 LGYFESAGYHCEAYNNPADFLLDINGD-----STAVALNREEDFKATEII 319

Qy 349 APCPPCPPEVDPI-----ESH-----TATSTLTQFCI 376
Db 320 EPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKITVFKETISYTSFCHOLRW 379

Qy 377 LFKRTFLSILRDVTLTHLRFMSHWVIGVLGLLYLHIGDDASKVFNNTGCLFFSMLFLMF 436
Db 380 VSKRSFKNLGNPQASQAIIIVTVGLVIGAIYFGLKNDSTGIONRAGVLFLLTNNQCF 439

Qy 437 AALMPTVLTTPLEMAVFMREHLNWSLKAYYLAKTMAV-VPEQVGVCPVVCISIVYMTG 495
Db 440 SSVSAVEL-FVVEKKLFIHEIYISGYRVSSYFLGKLLSLLPMRLPSIIFTICIVYMLG 498

Qy 496 QPAETSRFLFSALATATATLVAOSLGLLIGAANSLOVATFVGVPVTAIPVLLFSFFVSF 555
Db 499 LKPKADAFFVMMFTLMVAVSASSMALAIAAGOSVSVATLLMTICFVFMIFSGLLVNL 558
Qy 556 KTIPTYLOWSSYLSYRYGEGVILTYI-----GME-RGDLTCLCEERCPFRPQSILR 607
Db 559 TTIASWLSWLQYFESIPRYGFTALQHNFLQGNFCPGLNATGNNPCNYATCT-GEEYLKQ 617
Qy 608 ALQVEDAKLYMDFLVLGIFFLALRLAYLRLVRVK 643
Db 618 GIDLSPWGLKKNHVALACMIVIFLTIAYLKLLFLKK 653

RESULT 15
US-10-090-455-5
; Sequence 5, Application US/10090455
; Publication No. US20030027259A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hongyun
; APPLICANT: Le Bihan, Stephane
; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF
; FILE REFERENCE: 100103.406
; CURRENT APPLICATION NUMBER: US/10/090.455
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-090-455-5

Query Match 21.0%; Score 704; DB 9; Length 655;
Best Local Similarity 27.7%; Pred. No. 7.4e-55;
Matches 176; Conservative 141; Mismatches 247; Indels 72; Gaps 16;

Qy 61 IEVELSYV--REG--PCWRKRGYKTLKCLSGKFCRRELIGTMPSGAGKSTFMNLA 116
Db 37 LSFHNICRYVKLSGFLPC-RKPEKEILSNING-IMKPGNALIGPTGGKSSLLDVL 94

Qy 117 GYRE-SGMKQILVNGRPRELRTFRKMSCYIMODMLLPHLTVLEAMVMSANL--TEN 173
Db 95 ARKDPSGLSGDLVINGAPRP-ANFKCNSGVYVQDDVVMGTLTVRENLFQFSAALRLATTMT 153

Qy 174 PDVKNLVTLEILALGLMSCSHRTAL-----LSGGORKRLATALELVNPPVFFDEPT 228
Db 154 NHEKNERINRVIELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILSDEPT 213

Qy 229 SGDSASCFVSVLSKSLAOGGRTIITIHQPSAKLFEMEDKLYLSOGOCIFKGVVNTL 288
Db 214 TGLDSSANAVLLLLKRMKSQGRITIFSIHQPRYSIFKLFDSLTLASGRMLFHGPAQEA 273

Qy 289 IPYKGLGLHCPVYHNPADFIIEVASGEYGDLPNMFRAVONGLCMAEKKSSPEKNEVP 348
Db 274 LGYFESAGYHCEAYNNPADFLLDINGD-----STAVALNREEDFKATEII 319

Qy 349 APCPPCPPEVDPI-----ESH-----TATSTLTQFCI 376
Db 320 EPSKQDKPLIEKLAIEIVNSSFYKETAELHQLSGGKKKITVFKETISYTSFCHOLRW 379

Qy 377 LFKRTFLSILRDVTLTHLRFMSHWVIGVLGLLYLHIGDDASKVFNNTGCLFFSMLFLMF 436
Db 380 VSKRSFKNLGNPQASQAIIIVTVGLVIGAIYFGLKNDSTGIONRAGVLFLLTNNQCF 439

Qy 437 AALMPTVLTTPLEMAVFMREHLNWSLKAYYLAKTMAV-VPEQVGVCPVVCISIVYMTG 495
Db 440 SSVSAVEL-FVVEKKLFIHEIYISGYRVSSYFLGKLLSLLPMRLPSIIFTICIVYMLG 498

Qy 496 QPAETSRFLFSALATATATLVAOSLGLLIGAANSLOVATFVGVPVTAIPVLLFSFFVSF 555
Db 499 LKPKADAFFVMMFTLMVAVSASSMALAIAAGOSVSVATLLMTICFVFMIFSGLLVNL 558

608 ALDVEDAKLYMDFLVLGIFFLALRLLAYLVLRVK 643
:
: : : : :
618 GIDLSPWGLWNHVALACMIVIFLTAYLKLFLKK 653

Completed: June 10, 2003, 16:53:11
51 secs

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodcock T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195 (2000).
DR EMBL: AF003580; AAF51131.1; -
DR FlyBase: FBgn0031515; CG9664.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
KW ATP-binding.
SQ SEQUENCE 609 AA; 68287 MW; 20F9C519FEAE836F CRC64;
Query Match 32.5%; Score 1086.5; DB 5; Length 609;
Best Local Similarity 36.9%; Pred. No. 2.6e-72;
Matches 238; Conservative 118; Mismatches 234; Indels 55; Gaps 10;
1 MAEKALEAVGCGLCGAVAMAVTLEDGAEPPVLTTHLKKVNHTEAQRFSLPKRSVD 60
1 MADNAVQAPNGLGP-----QKQKALE 22
61 IEFVELSVREGPCWKRKGKYLTKLCLSGKFCRRELIGMPGSGAGKSTFMNLAGYRE 120
23 LHFQSVSYSLKGA---TKGSTPIINEACGVFKSGRLTAIPLGSGAGKSTLNLALGFKL 78
121 SGMKGQILVNGPRELRFKMSYIMQDMLLPHLTLEAMVMSANLNTN--PDVKN 178
79 QGVTFQFLNGRPRIMFERSKMSAYIAQNFVNLNLTVEETLRVSTDLKMPSSSTAQEKQ 138
179 DLVTEILTALGMSCHTTRTALLSGGQRKRLALELVNPNVPMFDEPTSLGDSASCFQ 238
139 KIIDIIDILQSCERTLVNLSGEGHRLSIGIELVTNPIMFDEPTSLGDCVSGYQ 198
239 VYSLMKSQAQGGRTICTIHQPSAKLFEMDFKLYILSOQGCIFKGVVNTLPIYKLGGLH 298
199 VICHQLRLAHDGRIVVGVHQPGRSLFQLFDDVLVLAHGEVLYAGQREMLPTEAQSGHI 258
299 CPTYHNPADFILEVASGEVDLNPMLFRAVQNGCLMAEKKSPEKNEVPAPCPPEPV 358
259 CPOYNPADFALEVCSSSTTERCESLITQNM--MHSTASNVYKLQVDEETALIDVHK 315
359 DPTE-SHTFATSTL---TQCFILKRTSLSDTTLVTHLRPMHSHVIGVLGLGLYHTG 414
316 DALDLHLRGKQGVGFWTQSLVLLRRLHRSRDMFVQMRVLMVHVALLGVVWQIG 375
415 DDASKVFNNTGCLTFSMFLMFAALMPTVLTPELMAVEMREHLYWYSLKAYILAKTMA 474
376 GDAQKIVSNVSLFFVILFPVAGNAMPSTILLCQMSDAVFIREYNGWYSLGAYILSKVLA 435
475 DVPFQVQVPCVYCSIVYVMTGQPAETSRFLFSALATATALVAQSLGLLGAASLQVA 534
436 DLPLQLTCTPMFISIGYCYMTGQPPPEQFAMCWCVMVTAFTGHFVIGTAGSLF-TMQLA 494
535 TEVGPVTAIPVLFLSGFVSFTIPTYLQWSSYLVSVRYGFGVILTIYGMERGDLTC-- 592
495 IELVSAATIPFLFGGFFIRLNLWSFLPICOVSSFFRIYFGLMRAIYGYDRGELECYA 554
593 LEERCFFREPQSIILRALOVEDAKLYMDFLVLGFIFFLALRLAYLV 637
555 TSNFCYRTAEQFLKDFQMEGNEFDWDMAVLGI-FILILLALFAFF 598

RESULT 15

Q8SZF5

ID Q8SZF5

AC Q8SZF5;

DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE RE02452p.
GN BEST:CK02656.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY070914; AAL48536.1; -
SQ SEQUENCE 576 AA; 64057 MW; B04135CB85BC89E4 CRC64;
Query Match 32.3%; Score 1079.5; DB 5; Length 576;
Best Local Similarity 41.6%; Pred. No. 7.9e-72;
Matches 226; Conservative 84; Mismatches 132; Indels 101; Gaps 6;
46 EAQRFSLPKRSADVIEFVELSVREGPCWKRKGKYLTKLCLSGKFCRRELIGMPGSG 105
31 QPKTLQHLPKRPVADLAFHNLTYRKEG--NRNNAKTILKGVSGRLSGELTAIMPSSG 87
106 ACKSTFMNLAGYRESGKQILVNGRPRELRFKMSYIMQDMLLPHLTLEAMVMS 165
88 AGKSTLLNLISGYKTSSTEGSVTMGAERNLSAFKRLSAYIMQDNLHGNLTVAQEA 147
166 ANLNTE---NPDVKNLVTLEILTALGMSCHTTRTALLSGGQRKRLALELVNPNV 222
148 TNLKSKFSKPE-KHSMIDILLTLSEHYTTRNLSSGQKRLSLALELVSNPPI 206
223 FDEPTSLGDSASCFQVSVSLMKSQAQGGRTICTIHQPSAKLFEMDFKLYILSOQ 282
207 FDEPTSLGDSSTCFQCIIHLKMLAAGRTVICTIHQPSARLFEMDFQLYTLADGCV 266
283 GVVNTLPIYKLGGLHCHPTYNHPADFILEVASGEVDLNPMLFRAV----- 328
267 GSTKQLVPFLSTLNLECPSTYHNPASVIEVSCGEHGHTRKLVDAIDNSKROVRSA 326
329 ----QNGLCAMAEKSSPEKN----- 345
327 GLKARNLVKVNKAILDKNDASSVSGRKYEDNLTNLNGLNGWVNDIVKEGNSVALV 386
346 -----EVPAPCPPEVDPIESHTFATSTLTQCFILKRTSLSDTTLVTH 393
387 TTNREGDAMIDVEKSNCTTALLTEITSPERYPTSQHFQVWVVKRTLLFSYRDWTL 446
394 LRFEMHVVIGVLGLYHIGODASKVFNNTGCLFSLMFLMFAALMPTVLTPELMAV 453
447 LRLFAHLVGLFGLYDIGNDGAKVLSNLGFLFENMLFMYTSMITITLSPLEMPVL 506
454 MREHLNYSKAYILAKTMAVDPFQVQVPCVYCSIVYVMTGQPAETSRFLFSALAT 513
507 LKENRWYSLKAYILASVADLPK-----LFSALSTCR 541
514 ALV 516
542 SCI 544

Search completed: June 10, 2003, 16:46:08

Job time : 92 secs

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobaray C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong F.N., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2193(2000).
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AB003580; AAF51130.1; -
DR FlyBase; FBgn0031516; CG9663.
DR InterPro; IPR0031593; AAA_Atpase.
DR InterPro; IPR003439; ABC_transportr.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transportr; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 812 AA; 90203 MW; CF3739101F12F3E CRC64;
Query Match 34.8%; Score 1163.5; DB 5; Length 812;
Best Local Similarity 35.7%; Pred. No. 7.1e-78;
Matches 281; Conservative 119; Mismatches 235; Indels 153; Gaps 19;
QY 2 AKKALEAVCGGLPGAVAMAVT-----LEDGAEPPVLTTHLK-----KVENHI--- 44
DB 29 SETHLPSSRGSGTGGALPRPTTPVPSKLLSRKPPHLTLNIAKASHDAPLAETNTVLTST 88
45 -----TEAQR-----FSLPKRS--AVDI 61
89 SESVGTPLGATPVSSPANQSORQASQSDSRNTSSTATSGGSIFPHEQYTKTKKINI 148
QY 62 EFVELSYVREGPCWKRKGYKTLKCLSGKFCRRELIGIMPSGAGKSTFMNLAGYRES 121
DB 149 GFENIRYTKFGVFORET--KDVLMGLTGYFKSGELSAVIGPSGAGKSTLLNLSGYTY 206
QY 122 GKKGQILVNGRPRLRTRFKNSCIYIMODMLPLHPLTVLEAMVSNANLNTEN--PDVRND 179
DB 207 GTGDFRNGNRDRDLKAKSNVAFTRQDTSLQAFLSVKAEAFHAFKXIGTHTHSEKRE 266
QY 180 LYTEILTALGLMSCSHRTTALLSGGQRKLALALELVNPNPVMFDEPTSGLDSDASCFQV 239
DB 267 RYKCLLEALGYNENRHTTGLSGGQKRLALALELVNPNPVLLIDPTTGLDSTSNQL 326
QY 240 VSLMSKLAQGGRTIITCIHOPSAKIFEMFDKLYILSQGQCFKGVVTVNLIPYKGLGLHC 299
DB 327 INLLKKALEGRTVICTIHPSAFTAMFDHLYAIGCKCIYAGGAQNLLPLGALNLHC 386
QY 300 PYYHNPADF---IIEVASGEYGDNL-----PMLFRAVQNG-----LCAMA 336
DB 387 PESYNPADYCEFTVMEIATHDVTADNQLKLVALMDNGRNEYRQSKRTARVAOLAAM- 445
QY 337 EKK-----SSP-EKNEVPA-----PCPP----- 353
DB 446 -KKVDQLMAAGLITPVTPAVMSTSPVAFPMQGNTEKPLTPINELSSRWVDSOTAGIGNAG 504

QY 354 --CPP-----EVDPIESH-----TFATSTLTQFCILFKRTFLSILRDTVLTHER 395
DB 505 SCKPKKKKNSKPAIEIDP--SHLCKRQNIYATPFVQLSILLVTRFLLIWRDSSLTWTNR 562
QY 396 FNSHVIVGLVIGLILYHIGDDASKVFNNTGCLFFSMLFMFAALMPTVTTFPLEMAVPMR 455
DB 563 FAHLITGLTGLTYFGIGNDAQTILNIFRYLFTYIMFYCAFSGILVKFFLEFPFIVSR 622
QY 456 EHLNMYSLKAYYLAKTMADVPPQVVCVIVYVWGTGQPAETSRLFLFSALATATATL 515
DB 623 EHFNRWYSLARIYVAITLADLIQIICSALFIVPYLYMTQQQLLELWRFQMFLLIVPTL 682
QY 516 VAQSLGLLIGAASNSIQVATFVGPVTAIPVLLFSGFVSEFKTIPTYLQWSSYLSVYRGF 575
DB 683 VSQSIGLAVGAAL-SLKLSILGPFPCFLOFSGFLEKMDAPVFLRMWDFISFLKYSL 741
QY 576 EGVLITVGMERGDUITCLERCPFPREPQSIILRALDVEDAKLYMDFLVGLGIFFLALRLAY 635
DB 742 EGATWAIFGYDREPLACNELYCHLRHPOFILKSLDMANGNYTLALIFLFGVLVFLRLAF 801
QY 636 LVLYRVK 643
DB 802 YIMSFLR 809
RESULT 14
QYQVQ4 PRELIMINARY; PRT; 609 AA.
ID QYQVQ4
AC QYQVQ4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE CG9664 protein.
GN CG9664.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Ashburner M., Henderson S.N.,
RA George R.A., Lewis S.E., Richards S., Zhang Q., Chen L.X.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobaray C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RESULT 12
Q9VJP9
ID **Q9VJP9** **PRELIMINARY:** **PRT:** **620 AA.**
AC **Q9VJP9**
DT **01-MAY-2000** (TREMBLrel. 13, Created)
DT **01-MAY-2000** (TREMBLrel. 13, Last sequence update)
DT **01-MAY-2000** (TREMBLrel. 13, Last sequence update)
DT **01-MAY-2000** (TREMBLrel. 13, Last sequence update)
DE **CG3164** protein.
OS **BEST:CK02656** OR **CG3164**.
OS **Drosophila melanogaster** (Fruit fly).
OC **Eukaryota**; **Metazoa**; **Arthropoda**; **Tracheata**; **Hexapoda**; **Insecta**;
OC **Pterygota**; **Neoptera**; **Endopterygota**; **Diptera**; **Brachycera**; **Muscomorpha**;
OC **Ephydroidea**; **Drosophilidae**; **Drosophila**.
NCBI_TaxID=7227;
NCBI_TaxID=7227;
SEQUENCE FROM N.A.
STRAIN=BERKELEY;
MEDLINE=20196006; **PubMed=10731132**;
RA **Adams M.D.**, **Celniker S.E.**, **Holt R.A.**, **Evans C.A.**, **Gocayne J.D.**,
RA **Amanatides P.G.**, **Scherer S.E.**, **Li P.W.**, **Hoskins R.A.**, **Galle R.F.**,
RA **George R.A.**, **Lewis S.E.**, **Richards S.**, **Ashburner M.**, **Henderson S.N.**,
RA **Sutton G.G.**, **Wortman J.R.**, **Yandell M.D.**, **Zhang Q.**, **Chen L.X.**,
RA **Brandon R.C.**, **Rogers Y.H.C.**, **Blazej R.G.**, **Champe M.**, **Pfeiffer B.D.**,
RA **Wan K.H.**, **Doyle C.**, **Baxter E.G.**, **Helt G.**, **Nelson C.R.**, **Miklos G.L.G.**,
RA **Abriel J.F.**, **Agbayani A.**, **An H.-J.**, **Andrews-Pfannkoch C.**, **Baldwin D.**,
RA **Balieu R.M.**, **Basu A.**, **Baxendale J.**, **Bayraktaroglu L.**, **Beasley E.M.**,
RA **Beeson K.Y.**, **Benos P.V.**, **Berman B.P.**, **Bhandari D.**, **Bolshakov S.**,
RA **Borkova D.**, **Botchan M.R.**, **Bouck J.**, **Brokstein P.**, **Brottier P.**,
RA **Burtis K.C.**, **Busam D.A.**, **Butler H.**, **Cadieu E.**, **Center A.**, **Chandra I.**,
RA **Cherry J.M.**, **Cawley S.**, **Dahlke C.**, **Davenport L.B.**, **Davies P.**,
RA **de Pablos B.**, **Delfech A.**, **Deng Z.**, **Mays A.D.**, **Dew I.**, **Dietz S.M.**,
RA **Dodson K.**, **Doup L.E.**, **Downes M.**, **Dugan-Rocha S.**, **Dunkov B.C.**, **Dunn P.**,
RA **Durbin K.J.**, **Evangeliasta A.C.**, **Ferraz C.**, **Fertiera S.**, **Fleischmann W.**,
RA **Foster K.**, **Gabrielian A.E.**, **Garg N.S.**, **Gelbart W.M.**, **Glasser K.**,
RA **Glodek A.**, **Gong F.**, **Grorrell J.H.**, **Gu Z.**, **Guan P.**, **Harris M.**,
RA **Harris N.L.**, **Harvey D.**, **Heiman T.J.**, **Hernandez J.R.**, **Houck J.**,
RA **Hostin D.**, **Houston K.A.**, **Howland T.J.**, **Wei M.-H.**, **Ibegwam C.**,
RA **Jalali M.**, **Kalush F.**, **Karpen G.H.**, **Ke Z.**, **Kennison J.A.**, **Ketchum K.A.**,
RA **Kimmel B.E.**, **Kodira C.D.**, **Kraft C.**, **Kravitz S.**, **Kulp D.**, **Lai Z.**,
RA **Lasko P.**, **Lei Y.**, **Levitky A.A.**, **Li J.**, **Li Z.**, **Liang Y.**, **Lin X.**,
RA **Liu X.**, **Mattai B.**, **McIntosh T.C.**, **McLeod M.P.**, **McPherson D.**,
RA **Merkulov G.**, **Milshina N.V.**, **Moharri C.**, **Morris J.**, **Moshrefi A.**,
RA **Mount S.M.**, **Moy M.**, **Murphy B.**, **Murphy L.**, **Muzny D.M.**, **Nelson D.L.**,
RA **Nelson D.R.**, **Nelson K.A.**, **Nixon K.**, **Nusskern D.R.**, **Pacleb J.M.**,
RA **Palazzolo M.**, **Pittman G.S.**, **Pan S.**, **Pollard J.**, **Puri V.**, **Reese M.G.**,
RA **Reinert K.**, **Remington K.**, **Saunders R.D.C.**, **Scheeler F.**, **Shen H.**,
RA **Shue B.C.**, **Siden-Kiamos I.**, **Simpson M.**, **Skupski M.P.**, **Smith T.**,
RA **Spier E.**, **Spadling A.C.**, **Stapleton M.**, **Strong R.**, **Sun E.**,
RA **Svirskas R.**, **Tector C.**, **Turner R.R.**, **Venter E.**, **Wang A.H.**, **Wang X.**,
RA **Wang Z.-Y.**, **Wassarman D.A.**, **Weinstock G.M.**, **Weissenbach J.**,
RA **Williams S.M.**, **Woodage T.**, **Worley K.C.**, **Wu D.**, **Yang S.**, **Yao Q.A.**,
RA **Ye J.**, **Yeh R.F.**, **Zaveri J.S.**, **Zhan M.**, **Zhang G.**, **Zhao Q.**, **Zheng L.**,
RA **Zheng X.H.**, **Zhong F.**, **Zhong W.**, **Zhou X.**, **Zhu S.**, **Zhu X.**, **Smith H.O.**,
RA **Gibbs R.A.**, **Myers E.W.**, **Rubin G.M.**, **Venter J.C.**;
RT "The genome sequence of *Drosophila melanogaster*."
RL **Science 287:2185-2195(2000)**.
CC -!- **SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.**
DR **EMBL: AE003590**; **AAFS1548.1**.
DR **FlyBase: FBgn0025683**; **BEST:CK02656**.
DR **InterPro: IPR003439**; **ABC_transporter**.
DR **Tram: PR00005**; **ABC_tran**; **1**.
DR **ProDom: PD000006**; **ABC_transporter**; **1**.
DR **ProSITE: PS00211**; **ABC_TRANSPORTER**; **1**.
KW **ATP-binding**; **Transport**.
SQ **SEQUENCE 620 AA: 70891 MW; 4EA5A336483E7BC5 CRC64;**
Query Match 37.1%; Score 1241.5; DB 5; Length 620;
Best Local Similarity 40.6%; Pred. No. 8.5e-84;
Matches 261; Conservative 103; Mismatches 184; Indels 95; Gaps 11;
QY 46 EAQRSHLPKRSADVIEFVELSYSVREG-----PCWRKRGYKTLKCLSGKF 92
DB 31 QPKTLQHLKRPVADLAFNLHTYPPVVRKGRNHNHSEGROWTPVVRHNTGICGQW-EY 89

[illegible]


```
Db 266 ELETIGLKVNTLTCNSGQKRLSIALELVNPNVMEFDEPTSGDSTCFQLISL 325
QY 243 MKSLAQGRTTICTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLPIYLKGLGHCTY 302
Db 326 LRSARGRTTIVCTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLPIYLKGLGHCTY 385
QY 303 HNPADFIIEVASGEVDLNPMLFRAVQNGLC----- 333
Db 386 HNPADYVLEVASGEDAVPKLVDAVKSACKYAHKDYVLTLAQGCNNDIIKSGSGA 445
QY 334 --AMA-----EKKSPKNEVPAPCC--PPEVD----- 360
Db 446 ENAMAILTEDEKPPLEDROLESPTVEDPADVPKPLETOQSONSDCSVVMNPTNAVDD 505
QY 361 -----IESHT-----PATSTLTQFCILF 378
Db 506 SCSFSSSGTQNAVGGSGGSPSVAVGWMTSLDSDSHSVTLPNKGTGPTSGWTFWILL 565
QY 379 KRTFSLRLDVLTHLRFMSHWVIGVLIGLLYHLIGDASKVFNNTGCLFMSLFLMFAA 438
Db 566 KRSFRTLRLDMLTHLRFMSHWVIGVIGAIIGMIYDVGNEASKIMSAGCIFVSLFTTFA 625
QY 439 LMPVTLTTPLENAVPMRHLNWTSLKAYILAKTMADVPFQVVCVVYCSIVYWMGTGPA 498
Db 626 MNPVTLTTPLENAVPMRHLNWTSLKAYILAKTMADVPFQVVCVVYCSIVYWMGTGPA 685
QY 499 ETSRFLFSALATATATVAQSLGILLIGAASNSLOVATFVGPVTAIPVLLFSGFSPKTI 558
Db 686 ELERSVMEVLVLSNLVAQSLGILLIGAGMN--IEFGVLPVPTTIPTILFSGFFVNDTI 744
QY 559 PTLQWSSYLVRYRGFEGVILTYGMERGDLTCLERCPREPOSILRALDVEDAKLYM 618
Db 745 PGLQWVYVSVRYRGFEGVILTYGMERGDLTCLERCPREPOSILRALDVEDAKLYM 804
QY 619 DELVL-GIFFALRLAVLVRYR 642
Db 805 DAVALIGIFF-ALRIAYFVLRWL 828

RESULT 9
Q8R055 PRELIMINARY: PRT: 335 AA.
ID Q8R055
AC Q8R055
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 37.2 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC026477; AAH26477.1;
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 335 AA; 37230 MW; CD3AD5FB0D3D9EDD CRC64;

Query Match 40.8%; Score 1364; DB 11; Length 335;
Best Local Similarity 80.4%; Pred. No. 3.4e-93;
Matches 267; Conservative 13; Mismatches 34; Indels 18; Gaps 3;

QY 196 TTTALLSGQKRLAIALELVNPNVPMV---FFDEPT-----SGLDSASCFQVYSL 242
Db 2 TKTNOFRGQGAFFLPVLSQKPPRLTHPLAISTSLVPGDLSGLDSASCFQVYSL 61
QY 243 MKSLAQGRTTICTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLPIYLKGLGHCTY 302
Db 62 MKSLAHGRTVICTTHQPSAKLFEMFDKLYILSQGQCIFKGVVNLPIYLKGLGHCTY 121
```

```
QY 303 HNPADFIIEVASGEVDLNPMLFRAVQNGLCMAEKSSPEKNEVPAPCCPPEVDPIE 362
Db 122 HNPADFIIEVASGEVDLNPMLFRAVQNGLCMAEKSSPEKNEVPAPCCPPEVDPIE 181
QY 363 SHTFATSTLTQFCILFRKRTFSLILRDVTLTHLRFMSHWVIGVLIGLLYHLIGDASKYFN 422
Db 182 SHTFATSTLTQFCILFRKRTFSLILRDVTLTHLRFMSHWVIGVLIGLLYHLIGDASKYFN 241
QY 423 NTGCLFSSMLFLMFAALMPTVLTTPLENAVPMRHLNWTSLKAYILAKTMADVPFQVVC 482
Db 242 NTGCLFSSMLFLMFAALMPTVLTTPLENAVPMRHLNWTSLKAYILAKTMADVPFQVVC 301
QY 483 PVVYCSIVYWMGTG-----QPAETSRFLFSAL 509
Db 302 PVVYCSIVYWMGTG-----QPAETSRFLFSAL 333

RESULT 10
Q9VL61 PRELIMINARY: PRT: 689 AA.
ID Q9VL61
AC Q9VL61
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CG5853 protein.
GN CG5853.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jafarli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
```

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sinden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 "The genome sequence of *Drosophila melanogaster*.";
 Science 287:2185-2195(2000).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AE003576; AAF51027.1; -.
 DR FlyBase: FBgn020762; Atet.
 DR InterPro: IPR003593; AAA_Atpase.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transportr; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 832 AA; 90965 MW; D32899B11805B64F CRC64;
 Query Match 51.5%; Score 1723; DB 5; Length 832;
 Best Local Similarity 47.1%; Pred. No. 2.4e-119;
 Matches 351; Conservative 109; Mismatches 153; Indels 132; Gaps 12;
 QY 10 GCGLPGAVAMAVTLEDGAEPPVLTTHLKKVENHITEAQR-----FSLPKRSAYDIEFV 64
 DB 104 GLGSNNHALAPKAVNNSGSP-----NQKKGTIALSHLPQRPVDFEFC 148
 QY 65 ELSYSVRGCPWRKRGYTKLKCSGKRCRELIGIMGPSGAGKSTFNILAGYRESOMK 124
 DB 149 DISYVSDS---HRRGFITLKSCKFRNGEITAIMGPSGAGKSTLNLNLAGYKTAOLS 205
 QY 125 GQILVNGPRELRFKMKSCYIMODMLLPHLTVLEAMVSNANLNTEN--PDVKNLDT 182
 DB 206 GSVLINSKERNLRERKLCYIMQDDVLNLTVREAMVAAKLGKGNMISYAKVVVVE 265
 QY 183 EILTALGLMSCSHTRTALLSGGQRKRLAIALELVNPPVPMFDEPTSGLDSASCFQVYSL 242
 DB 266 EILETIGLKVESVNTLTCNLSSGQRKRLSIALELVNPPVPMFDEPTSGLDSSTCFQLISL 325
 QY 243 MKSLAQGRTTICTHQPASAKLFEMFDKLYLSQGCQIFKGVVNLIPYKGLGHCPTY 302
 DB 326 LRLSARGGRTIVCTTHQPASAKLFEMFDKLYLSQGCQIFKGVVNLIPYKGLGHCPTY 385
 QY 303 HNPADFILEVASGEYGLDNLPMFLAVONGLC----- 333
 DB 386 HNPADYVLEVASGEYGLDNLPMFLAVONGLC----- 360
 QY 334 --AMA-----EKKSPKEVNPAPCP--PPEVDP----- 360
 DB 446 ENAMAILTDEKPPLEDRQLEPSIPVDDPAELKPKPLETOOSQNSDCSVNMPNTAVDD 505
 QY 361 -----IESHT----- 565
 DB 506 SCFSSSSKGTQNAVGGSGGSPSAVVGCMTSLDSSHESVVTLPNKTGFTSGTQFWILL 565
 QY 379 KRTFLSILRDVTLHREMSHWVIGVLIGLGLYHLHGDDASKVENNTGCLFFSMLFLMFAA 438
 DB 566 KRSEFTILRDKNLHRLFSHVIGVIGAILGMIIYDVGNASKMSNAGCFFVSLFTTTTA 625

QY 439 LMPTVLTFTPLEMAVFMREHLNWNYSKAYYLAKTMADVPFQVQVVPVYCSIVYMTGQPA 498
 DB 626 WMPTILTPTPTMSVFEHLNWNYSKAYYFAKTMADMPQIVFSSVYVLYVYLTSPQM 685
 QY 439 ETSRELLFSALATATATNALQAQSLGILLIGAASNLQVATFVGPVTAIPVLLFSGFFVSKTI 558
 DB 686 ELERYSMFVLCVLSNLSVAQSLGILLIGAAGMN-IETGVFLGPVTTTPTILFSGFFVNFDTI 744
 QY 559 PLYLOWSVLSYVRVYGFEGVILTYIGMBRGDLTCLCEERCPREPOSILRALDVEDAKLYM 618
 DB 745 PGYLOWTVTSYVRVYGFEGVIAIYGMORAKMQCQNMCHYVRPKRFLSEMSMDNALEFW 804
 QY 619 DFLVL-GIFFLALRLAYLVLRV 642
 DB 805 DAVALIGIFF-ALRIIAVFLRWKL 838
 ID Q9UAF0 PRELIMINARY; PRT: 832 AA.
 AC Q9UAF0;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE ABC transporter.
 GN ATET OR CG2969.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CANTON-S; TISSUE=TRACHEA;
 RA Yamamoto D., Jun1 N.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CANTON-S; TISSUE=TRACHEA;
 RX MEDLINE=97107435; PubMed=8950175;
 RA Kuwana H., Shimizu-Nishikawa K., Iwabata H., Yamamoto D.;
 RT "Molecular cloning and characterization of the ABC transporter
 RT expressed in *Trachea* (ATET) gene from *Drosophila melanogaster*.";
 RL Biochim. Biophys. Acta 1309:47-52(1996).
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL: AB030450; BAA83106.1; -.
 DR FlyBase: FBgn020762; Atet.
 DR InterPro: IPR003593; AAA_Atpase.
 DR Pfam: PF00005; ABC_tran; 1.
 DR ProDom: PD000006; ABC_transportr; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport.
 SQ SEQUENCE 832 AA; 90975 MW; 8F72ELC028AC06A3 CRC64;
 Query Match 51.5%; Score 1723; DB 5; Length 832;
 Best Local Similarity 47.1%; Pred. No. 2.4e-119;
 Matches 351; Conservative 109; Mismatches 153; Indels 132; Gaps 12;
 QY 10 GCGLPGAVAMAVTLEDGAEPPVLTTHLKKVENHITEAQR-----FSLPKRSAYDIEFV 64
 DB 104 GLGSNNHALAPKAVNNSGSP-----NQKKGTIALSHLPQRPVDFEFC 148
 QY 65 ELSYSVRGCPWRKRGYTKLKCSGKRCRELIGIMGPSGAGKSTFNILAGYRESOMK 124
 DB 149 DISYVSDS---HRRGFITLKSCKFRNGEITAIMGPSGAGKSTLNLNLAGYKTAOLS 205
 QY 125 GQILVNGPRELRFKMKSCYIMODMLLPHLTVLEAMVSNANLNTEN--PDVKNLDT 182
 DB 206 GSVLINSKERNLRERKLCYIMQDDVLNLTVREAMVAAKLGKGNMISYAKVVVVE 265
 QY 183 EILTALGLMSCSHTRTALLSGGQRKRLAIALELVNPPVPMFDEPTSGLDSASCFQVYSL 242

```

RESULT 5
Q9NT30 PRELIMINARY; PRT; 416 AA.
AC Q9NT30;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Hypothetical 46.6 kDa protein (Fragment).
GN DKF2P434P1420.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ottenwaelder B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137563; CAB70814.1;
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 416 AA; 46574 MW; 6FA5984FEA38A1E6 CRC64;
Query Match 64.6%; Score 2161; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 3.7e-152;
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 LDSASCFQVSVLMKSLAOGGRTIICTHQPSAKLFEMFDKLYTSLGQCIFKGVVNTLIP 290
DB 1 LDSASCFQVSVLMKSLAOGGRTIICTHQPSAKLFEMFDKLYTSLGQCIFKGVVNTLIP 60
QY 291 YLKGGLHCPYHNPADEIIEVASGEYGLNPMPLFRAVQNGLCMAEKKSSPEKNEVPAP 350
DB 61 YLKGGLHCPYHNPADEIIEVASGEYGLNPMPLFRAVQNGLCMAEKKSSPEKNEVPAP 120
QY 351 CPCPCPEVDPIESHFATSLTQFCILFKRTLSILRDVTLHLRPMHSHVIGLGLY 410
DB 121 CPCPCPEVDPIESHFATSLTQFCILFKRTLSILRDVTLHLRPMHSHVIGLGLY 180
QY 411 LHIGDDASKVFNTGCLFESMLFMPFAALMPTVLTFFLEMAVFMREHLNWSLKAYILA 470
DB 181 LHIGDDASKVFNTGCLFESMLFMPFAALMPTVLTFFLEMAVFMREHLNWSLKAYILA 240
QY 471 KTMADVPFQVCPVYVCSIVYWMTGQPAETSRFLFSALATATATVAQSLGLLIGAANS 530
DB 241 KTMADVPFQVCPVYVCSIVYWMTGQPAETSRFLFSALATATATVAQSLGLLIGAANS 300
QY 531 LQVATFVGPTAIPVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTIYGMERGDL 590
DB 301 LQVATFVGPTAIPVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTIYGMERGDL 360
QY 591 TCLEERCPEFQPSILRALDVEDAKLYMDFLVIGIFFLALRLAYLVLRVYKSER 646
DB 361 TCLEERCPEFQPSILRALDVEDAKLYMDFLVIGIFFLALRLAYLVLRVYKSER 416

RESULT 6
Q8RIU5 PRELIMINARY; PRT; 423 AA.
AC Q8RIU5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 47.5 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EVE;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; BC023077; AAH23077.1;
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 423 AA; 47464 MW; 9F3F2FB5B5117FDD CRC64;
Query Match 61.7%; Score 2064; DB 11; Length 423;
Best Local Similarity 94.7%; Pred. No. 5.8e-145;
Matches 396; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 229 SGLDSASCFQVSVLMKSLAOGGRTIICTHQPSAKLFEMFDKLYTSLGQCIFKGVVNTL 288
DB 6 SGLDSASCFQVSVLMKSLAOGGRTIICTHQPSAKLFEMFDKLYTSLGQCIFKGVVNTL 65
QY 289 IPLYKGLGLHCPYHNPADEIIEVASGEYGLNPMPLFRAVQNGLCMAEKKSSPEKNEVP 348
DB 66 IPLYKGLGLHCPYHNPADEIIEVASGEYGLNPMPLFRAVQNGLCMAEKKSSPGKNELP 125
QY 349 APCPCPEVDPIESHFATSLTQFCILFKRTLSILRDVTLHLRPMHSHVIGLGL 408
DB 126 AHCPTCPPELDPIESHFATSLTQFCILFKRTLSILRDVTLHLRPMHSHVIGLGL 185
QY 409 LYLHIGDDASKVFNTGCLFESMLFMPFAALMPTVLTFFLEMAVFMREHLNWSLKAY 468
DB 186 LYLHIGDDASKVFNTGCLFESMLFMPFAALMPTVLTFFLEMAVFMREHLNWSLKAY 245
QY 469 LAKTMADVPFQVCPVYVCSIVYWMTGQPAETSRFLFSALATATATVAQSLGLLIGAAS 528
DB 246 LAKTMADVPFQVCPVYVCSIVYWMTGQPAETSRFLFSALATATATVAQSLGLLIGAAS 305
QY 529 NSLQVATFVGPTAIPVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTIYGMERG 588
DB 306 TSLQVATFVGPTAIPVLLFSGFFVSKTIPTIYLOWSSYLSYVRYGFEVILTIYGMERG 365
QY 589 DLTCLERCPEFQPSILRALDVEDAKLYMDFLVIGIFFLALRLAYLVLRVYKSER 646
DB 366 HLTCLDEQCPFRDPQIILRELDEVEAKLYMDFLVIGIFFLALRLAYLVLRVYKSER 423

RESULT 7
Q9VQY4 PRELIMINARY; PRT; 832 AA.
AC Q9VQY4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ATET protein.
GN ATET OR CG2969.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```

```

SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RA Oldfield S., Lowry C.A., Lightman S.L.;
RT "Cloning and expression of a novel mammalian white family ABC-
RT transporter:white2.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AJ303374; CAC2156.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR005284; Pigment_permease.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR SMART: SM00382; AAA; 1.
DR TIGRFAMs: TIGR00955; 3a01204; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding.
KW
SO SEQUENCE 666 AA; 74146 MW; 67298A0D99EE0409 CRC64;

Query Match
Best local similarity 70.5%; Score 2360; DB 11; Length 666;
Matches 453; Conservative 79; Mismatches 89; Indels 4; Gaps 2;

QY 26 DGAEPVPLVTHLKVENHIETAQRFSHLPKRSAVDIEFVELSVYRGPCWRKRGYKTL 85
DB 42 DELETEDLLNGHLKKVDNNFTAEQRFSSLPRAAVNIEFKDLSVYSGVPWRKKGKYL 101
QY 86 KCLSGKFCRRELICIMGSPCAGKSTFNWILAGYRESGMKGQILVNGRPRLRFRKMVSC 145
DB 102 KGISGFENSELVAIMGSPCAGKSTLNWILAGYRETGMKGAVLINGMRDLRCFRKVSC 161
QY 146 IMQDDMLLPHLTVLEAMVVSANLNLTENPOVKNDLVTEILTALGLMSCSHTRTALLSGG 205
DB 162 IMQDDMLLPHLTVOEAMVVSANLKLQERDEGRREVMYKEILTALGLPCANTRTGSLSGG 221
QY 206 RKRLAIALELVNPPVNFDEPTSGLDSASCFQVWSLMKSLAQGGRITICTIHQPSAKLF 265
DB 222 RKRLAIALELVNPPVNFDEPTSGLDSASCFQVWSLMKSLAQGGRISIVCTIHQPSAKLF 281
QY 266 EMFDKLYLSOGOCIFKGVYTNLIPYLGKGLGHCPTYHNPADFTIEVASGEYGLNPMFL 325
DB 282 ELFDQLYVLSOGQCVYRGKVSNLVPLVLRDLGLNCPPTHNPADFVMEVASGEYGDQNSRLV 341
QY 326 RAYQVNGLCAMAEKKSSPEKNEVPACPCPPPEVDPIE--SHTFATSLTQFCILFKRTF 382
DB 342 RAVREGMCDSDYKRELGGDGVNPFMLWHRPAEEDSASMEGCHSFASCLTQFCILFKRTF 401
QY 383 LSLRDTVTLTHLRFSMSHVTVGLIGLLYLHIGDDASKVFVNTGCLFSPMLFLMFAALMPT 442
DB 402 LSLMRDVLTHLRITSHIGLIGLLYLHIGNEAKKVLNSGSLFFTSMLFLMFAALMPT 461
QY 443 VLFPPLEMAVFMREHLNLYWSLKAYLAKTMADVPOVYCPVYVCSIVYVMTGQPAETSR 502
DB 462 VLFPPLEMSVFLREHLNLYWSLKAYLAKTMADVPOIMEFPVAYCSIVYVMTSQPSDAVR 521
QY 503 FLLESALATATAVLAQSLGLLIGAASNSLOVATFVGVTVAIPVLLFGSFVFSKTIPTYL 562
DB 522 FVLFAALGTWTSVLAQSLGLLIGAASNSLOVATFVGVTVAIPVLLFGSFVFSKTIPTYL 581
QY 563 QWSSYLSYVRYGEGVILTYTYGMERGDLTC-LERCPFPQSIILRALDVEDAKLYMDFL 621
DB 582 QWMSYLSYVRYGEGVILSYTYGDLREDLHCDIAETCHFOKSEAILRELDVENAKLYLDFI 641
QY 622 VLGIFFLALRLLAIVLRYRVKSPER 646
DB 642 VLGIFFLSLRLAIYVFLRYKRAER 666

RESULT 4
Q96L76 PRELIMINARY; PRT: 785 AA.
ID Q96L76
AC Q96L76;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

```

DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	ATP-binding cassette transporter G1.
GN	ABCG1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RS	SEQUENCE FROM N.A.
RP	MEDLINE=21486468; PubMed=11500512;
RX	Kennedy M.A., Venkateswaran A., Tarr P.T., Xenarios I., Kudoh J.,
RA	Shimizu N., Edwards P.A.,
RT	"Characterization of the human ABCG1 gene. Liver X receptor activates
RT	an internal promoter that produces a novel transcript encoding an
RT	alternative form of the protein."
RL	J. Biol. Chem. 276:39438-39447(2001).
DR	EMBL; AY048757; AACL06598.1; ..
DR	InterPro; IPR003439; ABC_transportr.
DR	InterPro; IPR003603; LRRcap.
DR	InterPro; IPR005284; Pigment_permease.
DR	Pfam; PF00005; ABC_tran; 1.
DR	ProDom; PD000006; ABC_transportr; 1.
DR	SMART; SM00446; LRRcap; 1.
DR	TIGRFAMS; TIGR00955; 3a01204; 1.
DR	PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW	ATP-binding
QY	SEQUENCE 785 AA; 86623 MW; 7F06DC1F60F26A44 CRC64;
Query Match 65.0%; Score 2177; DB 4; Length 785;	
Best Local Similarity 70.5%; Pred. No. 5.1e-153;	
Matches 420; Conservative 71; Mismatches 67; Indels 38; Gaps 4;	
Qy	78 KRCYKTLKCLGSKFCRRELIGIMPGSGACKSTFNNILAGYRESGMKGQIILVNGRPRLR 137
Db	201 EEGTKTLKIGSKFNGSELVIMPGSGAKSTLNNILAGYRGTGMKGAVLINGLPDLR 260
Qy	138 TFRKMSCYIMQDDMLLPHLTVLEAMMYSANLINTENPDVKNDLVTEIITALGLMSCSHTR 197
Db	261 CFRKVSICYIMQDDMLLPHLTVQEAAMMYSALHLKQEKDEGRREMYKEILTALGLLSCANTR 320
Qy	198 TALLSGQQRRLAIALELVNPPVWFDEPTSGLDSASCFQVVSMLKSLAOGRTIICIT 257
Db	321 TGLSSGQQRRLAIALELVNPPVWFDEPTSGLDSASCFQVVSMLKSLAOGGRSIICIT 380
Qy	258 HQPSAKLFEMFDKLYILSQCOCIFPKGVVNTLIPYLKGLGLHCPYHNPAADPIIEVASGEY 317
Db	381 HQPSAKLFEFLDQLYLSQCQCYYRGKVCNLVPYLRDLGLNCPYHNPAADPFVMEVASGEY 440
Qy	318 GDLNPLMFLRAVQNLCAAMEKKSPKEKNEVPAPCPPEVDPI----- 361
Db	441 GDQNSRLRAVRGCMC-----DSDKRDLDGGDA----EVPNPLWHRPSEVKQTKRLK 489
Qy	362 -----ESTHTATSTLTORCILPKRFTLSLTRDTVTLTRFMSHVTVGLIGLLYL 411
Db	490 GLRKDSSMEGCHSFSSASCITQCILPKRFTLSIMRDSVTLHLRITSHIGLIGLLDYL 549
Qy	412 HIGDASKVFNNTGCLFFSMLFMFAALMPTVTLTFPELMAFMVFMREHLNWNYSKAYYLAK 471
Db	550 GIGNETKVKVLSNGELFFSMLFMFAALMPTVTLTFPELMEGVFLREHLNWNYSKAYYLAK 609
Qy	472 TMADVPPQVVCVYCSIVVWMTGQPAETSRFLFLFSALATATALVAQSLGLLIGAA NSL 531
Db	610 TMADVPPQIMFPVAYCSIVVWMTSOPSDAVRFVLFAALGTMTSLVAQSLGLLIGAA NSTL 669
Qy	532 QVATFVGPVTAIPVLLFSGFEVFSFKTIPTVLQNSSYLSVYRYGEGVILTITTYGMERGDLT 591
Db	670 QVATEVGPVTAIPVLLFSGFEVSEDITPTYLQWMSYISVYRYGEGVILSYGLDREDH 729
Qy	592 C-LEERCPPREPOSILRALDVEDAKLYMDFLVUGIIFTLALRLIAYLVLYRVKSER 646
Db	730 CDIDECHFQKSAIRLELDEVAKLYLDFIVLIGIIFTSIRLIAYLVLYRVKIRAR 785

RECEIVED

```

DR Pfam: PF00005; ABC_tran: 1.
DR ProDom: PD000006; ABC_transportr: 1.
DR SMART: SM00382; AAA; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 646 AA: 72097 MW: ED834AC4301ED6FB CRC64;

Query Match 95.7%; Score 3202; DB 11; Length 646;
Best Local Similarity 95.5%; Pred. No. 4.2e-229;
Matches 617; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKVKVENHTEAQRFSHLPKRSAYD 60
DB 1 MAEKALEAVGCGLGPGAVAMAVTLEDGAEPPVLTTHLKVKVENHTEAQRFSHLPKRSAYD 60
QY 61 IEFVELSVREGPCWKRKGKTKLLKLSGKFCRRELIGIMPSGAGKSTFMNLAGYRE 120
DB 61 IEFVELSVREGPCWKRKGKTKLLKLSGKFCRRELIGIMPSGAGKSTFMNLAGYRE 120
QY 121 SGMKQIILVNGRPRELRTFRKMSCYIMODDMLLPHLTVALEAMVSNLNTENPDVKNDL 180
DB 121 SGMKQIILVNGRPRELRTFRKMSCYIMODDMLLPHLTVALEAMVSNLNTENPDVKNDL 180
QY 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPNPFDEPTSGLDASCFQV 240
DB 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPNPFDEPTSGLDASCFQV 240
QY 241 SLMKSIAOGGRTIICTIHQPSAKLFEMEDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCP 300
DB 241 SLMKSIAOGGRTIICTIHQPSAKLFEMEDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCP 300
QY 301 TYHNPADRIIEVASGEYGDLPMPFRAVQNGLCMAEKKSPKNEVPAPCCPPEYDP 360
DB 301 TYHNPADRIIEVASGEYGDLPMPFRAVQNGLCMAEKKSPKNEVPAPCCPPEYDP 360
QY 361 IESHTFATSTLTCILPKRFLSLTRDTVLTHLRFMSHVIGVLIIGLILYHIGDDASKV 420
DB 361 IESHTFATSTLTCILPKRFLSLTRDTVLTHLRFMSHVIGVLIIGLILYHIGDDASKV 420
QY 421 FNNTGCLFESMLFLMFAALMPTVTLFPLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQV 480
DB 421 FNNTGCLFESMLFLMFAALMPTVTLFPLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQV 480
QY 481 VCPVYVCIVVMTGQPAETSRFLFSALATATALVAQSLGILLGAASNSLQVATFVG 540
DB 481 VCPVYVCIVVMTGQPAETSRFLFSALATATALVAQSLGILLGAASNSLQVATFVG 540
QY 541 TAIPVLLSGFVSFKTIPTIYLOWSSYLSYVRYGEGVILTYGMERGDLTCLERCPR 600
DB 541 TAIPVLLSGFVSFKTIPTIYLOWSSYLSYVRYGEGVILTYGMERGDLTCLERCPR 600
DB 601 DPQIILRELDEVEAKLYMDFVLGIFFLALRLALLVLLRYRVKSER 646
DB 601 DPQIILRELDEVEAKLYMDFVLGIFFLALRLALLVLLRYRVKSER 646

RESULT 2
Q91WA9 PRELIMINARY; PRT: 627 AA.
AC Q91WA9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to ATP-binding cassette, sub-family G (White), member 4.
GN 6430517004RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_Taxid=10090;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE, AND RETINA;
RA Strausberg R.;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC016200; AAH16200.1;
DR MGD: MGI:1924137; 6430517004RIK.
DR InterPro: IPR003439; ABC_transportr.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding.
SQ SEQUENCE 627 AA: 70371 MW: 28684468CB125D64 CRC64;

Query Match 92.8%; Score 3106; DB 11; Length 627;
Best Local Similarity 95.4%; Pred. No. 5.3e-22;
Matches 598; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

QY 20 MAVTLEDGAEPPVLTTHLKVKVENHTEAQRFSHLPKRSAYDIEFVELSVREGPCWKR 79
DB 20 MAVTLEDGAEPPVLTTHLKVKVENHTEAQRFSHLPKRSAYDIEFVELSVREGPCWKR 79
QY 80 GYKTLKCLSGKFCRRELIGIMPSGAGKSTFMNLAGYRESGMKGQILVNGRPRELRT 139
DB 80 GYKTLKCLSGKFCRRELIGIMPSGAGKSTFMNLAGYRESGMKGQILVNGRPRELRT 139
QY 140 RKMSCYIMODDMLLPHLTVALEAMVSNLNTENPDVKNDLVTETLTALGLMSCSHRTA 199
DB 140 RKMSCYIMODDMLLPHLTVALEAMVSNLNTENPDVKNDLVTETLTALGLMSCSHRTA 199
QY 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPNPFDEPTSGLDASCFQV 259
DB 181 VTEILTALGLMSCSHTRTALLSGGORKRLATALELVNPNPFDEPTSGLDASCFQV 259
QY 200 LLSGQQRRLATALELVNPNPFDEPTSGLDASCFQVVSMLKSLAOGGRTIICTIHQ 299
DB 200 LLSGQQRRLATALELVNPNPFDEPTSGLDASCFQVVSMLKSLAOGGRTIICTIHQ 299
QY 260 PSAKLFEMFDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCPYHNPADRIIEVASGEYGD 319
DB 260 PSAKLFEMFDKLYILSQGOCIFKGVVNTLIPYLKGLGLHCPYHNPADRIIEVASGEYGD 319
QY 320 LNPMLFRAVQNGLCMAEKKSPKNEVPAPCCPPEVDPTESHTATSTLTQFCILFK 379
DB 320 LNPMLFRAVQNGLCMAEKKSPKNEVPAPCCPPEVDPTESHTATSTLTQFCILFK 379
QY 301 LNPMLFRAVQNGLCMAEKKSPKNEVPAPCCPPEVDPTESHTATSTLTQFCILFK 360
DB 301 LNPMLFRAVQNGLCMAEKKSPKNEVPAPCCPPEVDPTESHTATSTLTQFCILFK 360
QY 380 RTFLSILRDTVLTHLRFMSHVIGVLIIGLILYHIGDDASKVFNNTGCLFESMLFLMFAAL 439
DB 380 RTFLSILRDTVLTHLRFMSHVIGVLIIGLILYHIGDDASKVFNNTGCLFESMLFLMFAAL 439
QY 440 MPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQVVCVIVVMTGQPAE 499
DB 440 MPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQVVCVIVVMTGQPAE 499
QY 421 MPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQVVCVIVVMTGQPAE 480
DB 421 MPTVLTFFLEMAVFMREHLNLYWSLKAYYLAKTMADVPFQVVCVIVVMTGQPAE 480
QY 500 TSRLFLFSALATATALVAQSLGILLGAASNSLQVATFVGPTAIPVLLSGFVSFKTIPT 559
DB 500 TSRLFLFSALATATALVAQSLGILLGAASNSLQVATFVGPTAIPVLLSGFVSFKTIPT 559
QY 481 TSRLFLFSALATATALVAQSLGILLGAASNSLQVATFVGPTAIPVLLSGFVSFKTIPT 540
DB 481 TSRLFLFSALATATALVAQSLGILLGAASNSLQVATFVGPTAIPVLLSGFVSFKTIPT 540
QY 560 TYLOWSSYLSYVRYGEGVILTYGMERGDLTCLERCPRPQSIILRALDVEDAKLYMD 619
DB 560 TYLOWSSYLSYVRYGEGVILTYGMERGDLTCLERCPRPQSIILRALDVEDAKLYMD 619
QY 541 TYLOWSSYLSYVRYGEGVILTYGMERGDLTCLERCPRPQSIILRALDVEDAKLYMD 600
DB 541 TYLOWSSYLSYVRYGEGVILTYGMERGDLTCLERCPRPQSIILRALDVEDAKLYMD 600
QY 620 FLVIGIFFLALRLALLVLLRYRVKSER 646
DB 620 FLVIGIFFLALRLALLVLLRYRVKSER 646
QY 601 FLVIGIFFLALRLALLVLLRYRVKSER 627.
DB 601 FLVIGIFFLALRLALLVLLRYRVKSER 627.

RESULT 3
Q9EPG9 PRELIMINARY; PRT: 666 AA.
AC Q9EPG9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE ABC transporter, white homologue.
GN ABC8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_Taxid=10116;
RX [1]

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:16:04 ; Search time 87 Seconds
(without alignments)
1529.959 Million cell updates/sec

Title: US-10-072-621-9

Perfect score: 3347

Sequence: 1 MAEKALEAVGGLPGGAVAM.....FLALRLLAYLVLYRVKRSR 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

rched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

(SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3202	95.7	646	11 Q8VBS9	Q8vbs9 mus musculus
2	3106	92.8	627	11 Q91WA9	Q91wa9 mus musculus
3	2360	70.5	666	11 Q9EPG9	Q9epg9 rattus norv
4	2177	65.0	785	4 Q96L76	Q96l76 homo sapien
5	2161	64.6	416	4 Q9NT30	Q9nt30 homo sapien
6	2064	61.7	423	11 Q8RIU5	Q8riu5 mus musculus
7	1723	51.5	832	5 Q9VQV4	Q9vqv4 drosophila
8	1723	51.5	832	5 Q9UAF0	Q9uaf0 drosophila
9	1364	40.8	335	11 Q8R0S5	Q8r0s5 mus musculus
10	1322	39.5	689	5 Q9VL61	Q9vl61 drosophila
11	1319	39.4	689	5 Q96OD7	Q96od7 drosophila
12	1241.5	37.1	620	5 Q9VPJ9	Q9vpj9 drosophila
13	1163.5	34.8	812	5 Q9VQN5	Q9vqn5 drosophila
14	1086.5	32.5	609	5 Q9VQN4	Q9vqn4 drosophila
15	1079.5	32.3	576	5 Q8SZF5	Q8szf5 drosophila
16	1070	32.0	615	5 Q8T998	Q8t998 drosophila

17	1067	31.9	615	5 Q9VQF1	Q9vqf1 drosophila
18	1060	31.7	766	5 Q8T9E6	Q8t9e6 drosophila
19	1051	31.4	788	5 Q9VQ41	Q9vq41 drosophila
20	1016.5	30.4	643	5 Q9VPJ7	Q9vpj7 drosophila
21	950.5	28.4	232	4 Q43576	Q43576 homo sapien
22	921.5	27.5	623	5 Q8SXX6	Q8sxx6 drosophila
23	888	26.5	597	5 Q9VTL3	Q9vtl3 drosophila
24	873.5	26.1	687	5 Q94960	Q94960 drosophila
25	845.5	25.3	687	5 Q9NH94	Q9nh94 bombyx mori
26	820	24.5	679	5 Q9BH97	Q9bh97 ceratitidis c
27	815.5	24.4	670	5 Q77423	Q77423 bactrocera
28	783	23.4	669	5 Q8WRE2	Q8wrf2 tribolium c
29	782.5	23.4	801	5 Q8T691	Q8t691 dictyosteli
30	777	23.2	669	5 Q8WRR1	Q8wrr1 tribolium c
31	770	23.0	672	10 Q9LI82	Q9li82 arabidopsis
32	767.5	22.9	648	10 Q9C6W5	Q9c6w5 arabidopsis
33	764.5	22.8	692	5 P91892	P91892 aedes aegypt
34	760.5	22.7	646	10 Q9C6R7	Q9c6r7 arabidopsis
35	719	21.5	657	11 Q9R004	Q9r004 mus musculu
36	717	21.4	655	4 Q96LD6	Q96ld6 homo sapien
37	716.5	21.4	658	5 Q16574	Q16574 caenorhabdi
38	712	21.3	655	4 Q96TA8	Q96ta8 homo sapien
39	705.5	21.1	662	10 Q949Y4	Q949y4 arabidopsis
40	699.5	20.9	609	10 Q9C8W6	Q9c8w6 arabidopsis
41	697	20.8	737	10 Q9FT51	Q9ft51 arabidopsis
42	691.5	20.7	751	10 Q93Y54	Q93y54 arabidopsis
43	689	20.6	668	10 Q9ARU4	Q9aru4 oryza sativ
44	686	20.5	695	5 Q19585	Q19585 caenorhabdi
45	685.5	20.5	1476	5 Q965D3	Q965d3 dictyosteli

ALIGNMENTS

RESULT 1

Q8VBS9 ID Q8VBS9 PRELIMINARY; PRT: 646 AA.
AC Q8VBS9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative white family ABC-transporter (ATP-binding cassette transporter sub-family G member 4) (ATP-binding cassette transporter ABCG4).
DE ABCG4).
GN ABCG4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Oldfield S., Lowry C.A., Lightman S.L.;
RT "Cloning and expression of a mammalian white family ABC-transporter, ABCG4".
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Annino T., Tammur J., Hutchinson A., Rzhetsky A., Dean M., Allikmets R.;
RT "Human and mouse orthologs of a new ATP-binding cassette gene, ABCG4".
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C; TISSUE=BRAIN;
RA Yoshikawa M., Yabuuchi H., Ikegami Y., Ishikawa T.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ426047; CAD19779.2;
DR EMBL; AF411084; AAL57369.1;
DR EMBL; AY040865; AAK91781.1;
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transportr.

Search completed: June 10, 2003, 16:47:52
Job time : 28 secs

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/772,270A
FILING DATE: December 23, 1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 6580-81
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 711 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE: Actinobacillus pleuropneumoniae
US-08-772-270A-12

Query Match 6.9%; Score 230.5; DB 3; Length 711;
Best Local Similarity 26.3%; Pred. No. 1.8e-14;
Matches 82; Conservative 54; Mismatches 103; Indels 73; Gaps 10;
QY 18 VAMAVTLEDGAPPPVLTTHLKKVENHTEAQR-----FSLHPKRSVNDIEFV 64
Db 429 VRLAQLWDFQFQVIGISITRLGDVLSPTENYQKLSLPEIFGDIAPKHFIRFYKPDAPIL 488
QY 65 --ELSVREGPCWRKRGYKTLKCLSGFCRELIGIMPGSGAGKSTFMNLAGYREGM 122
Db 489 LDDVNLVSKQG-----EVIGIVRSGSGKSTLTLLQRF-YIP 525
QY 123 MKGOILVNGRPREL---RFRKMSCYIMODDML-----PHLT---VLEAMM 163
Db 526 ENQOVLIDGHDLADPNWLRQIGVVDNVLNRSIRDNTALDPSMSERVYIAKL 585
QY 164 VSNANLNTENPDVKNLVTETLTALGLMSCSHTRTALLSGQKRLALALELVNPPVMF 223
Db 586 AGAHDFISELRECYNTIVGEL-----GAGLSGQQRQRTAARALVNNPRILI 632
QY 224 FDEPTSGLSASCFQVSVLMKSLAQGGRILCTIHQPSAKLFEMFDKLVLSQGCIFKG 283
Db 633 FDEATSAIDYSEHIIMQNMOKICH-GRTVIIAHLST--VKNADRIIVMEKGHIVEQG 689
284 VVTNLPYLYLKL 295
Db 690 KHNQLENENGL 701

RESULT 13
US-09-305-984-18
Sequence 18, Application US/09305984B
Patent No. 6331407
GENERAL INFORMATION:
APPLICANT: No. 6331407ak, Rodger
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016N1
CURRENT APPLICATION NUMBER: US/09/305,984B
CURRENT FILING DATE: 1999-05-05
EARLIER APPLICATION NUMBER: 60/084,399
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: 09/305,984
EARLIER FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 76
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 18
LENGTH: 215
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-305-984-18

Query Match 6.7%; Score 223; DB 4; Length 215;
Best Local Similarity 36.6%; Pred. No. 1.5e-14;
Matches 71; Conservative 30; Mismatches 61; Indels 32; Gaps 10;
QY 90 GKFCRELIGIMPGSGAGKSTFMNLAGYREGMGOILVNGRPREL---TFRKMS-- 143
Db 31 GKF-----YSIIGESGAGKSTLLSLLAGL-DSPVEGSILFQG--EDIRKKGYSYHRMHI 82
QY 144 CYIMODDMLPHLTVLEAMMVSNANLNTENPDVKNLVTETLTALGLMSCSHTRTAL-LS 202
Db 83 SLVFQNYNLDYLSPLE-----NIRLVNKKASKNTLL-----ELGLDSESQIKRNVQLQS 131
QY 203 GGQRRRLALALELVNPPVMFDEPTSGLDSCFQVSVLMKSLAQ-GGRTTICTITH--- 258
Db 132 GGQQRVAIARSIVSEAPVILADEPTGNLDPKTAGDIVELLKSLAQKTKCVIVVTHSKE 191
QY 259 ---OPSAKLFEMFDK 270
Db 192 VAQASDITLELKDK 205

RESULT 14
US-09-073-541A-18
Sequence 18, Application US/09073541A
Patent No. 6448224
GENERAL INFORMATION:
APPLICANT: No. 6448224ak, Rodger
APPLICANT: Toumanen, Elaine
TITLE OF INVENTION: NOVEL ANTIBIOTICS AND METHODS OF USING THE SAME
FILE REFERENCE: 1340-1-016
CURRENT APPLICATION NUMBER: US/09/073,541A
CURRENT FILING DATE: 1998-05-06
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 215
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-073-541A-18

Query Match 6.7%; Score 223; DB 4; Length 215;
Best Local Similarity 36.6%; Pred. No. 1.5e-14;
Matches 71; Conservative 30; Mismatches 61; Indels 32; Gaps 10;
QY 90 GKFCRELIGIMPGSGAGKSTFMNLAGYREGMGOILVNGRPREL---TFRKMS-- 143
Db 31 GKF-----YSIIGESGAGKSTLLSLLAGL-DSPVEGSILFQG--EDIRKKGYSYHRMHI 82
QY 144 CYIMODDMLPHLTVLEAMMVSNANLNTENPDVKNLVTETLTALGLMSCSHTRTAL-LS 202
Db 83 SLVFQNYNLDYLSPLE-----NIRLVNKKASKNTLL-----ELGLDSESQIKRNVQLQS 131
QY 203 GGQRRRLALALELVNPPVMFDEPTSGLDSCFQVSVLMKSLAQ-GGRTTICTITH--- 258
Db 132 GGQQRVAIARSIVSEAPVILADEPTGNLDPKTAGDIVELLKSLAQKTKCVIVVTHSKE 191
QY 259 ---OPSAKLFEMFDK 270
Db 192 VAQASDITLELKDK 205

RESULT 15
US-09-134-001C-4595
Sequence 4595, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al

```

QY      288 LIPYLK 293
      :      :
Db      229 IFNHPK 234

RESULT 10
US-09-134-001C-4042
; Sequence 4042, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4042
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4042

Query Match      7.0%; Score 233.5; DB 4; Length 341;
Best Local Similarity 27.9%; Pred. No. 2.7e-15;
Matches 74; Conservative 60; Mismatches 104; Indels 27; Gaps 9;

QY      61 IEFVELSYSVREGPCWRKRGYKTLKLSGKFCRRRELIGMGPSGAGKSTFMNLAGYRE 120
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      2 IEFKNVKNYFR-----KKRETQALKNVSKFDQHDIFGVIGYSGAGKSTLYR-LVNQLE 55

QY      121 SGMKGGIOLVNGRPRLRTRFKMSYIMQDD--MLPLHLTVLEAMVYSAN-----LNLTE 172
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      56 TVSDGQVIVDG--HEIDTYKEKDLRDKDKDIGMIFOHFNLLNSKSVYKNVAMPLILSKTN 113

QY      173 NPDKVNDLVEITLTALGLMSCSTRPALLSGGORKELATALELVNPPVPMFFDEPTSGILD 232
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      114 KEIK-EKVDLEMEFVGLADKKDQPPDELSGGQKQKVAIARALVTHPKILLCDKATFSLD 172

QY      233 SASCFOWSLMSLAQG-GRITICTHOPSAKLFEMFDKLYLSOGOCIFKGVVTNLIPY 291
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      173 PATTSILNLSNVNRTFCVTIMITHMSV-IQKICHRVAVMNGEVIEMGTVDKDFVSH 231

QY      292 LKGLGLHCPTYHNPADFIIEVASGE 316
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db      232 -----PQNTAKNFVSTVINTE 248

RESULT 11
US-08-402-804-6
; Sequence 6, Application US/08402804
; Patent No. 5874300
; GENERAL INFORMATION:
; APPLICANT: Blaser, Martin J.
; APPLICANT: Pei, Zhiheng
; TITLE OF INVENTION: Campylobacter Jejuni Antigens, And
; NUMBER OF INVENTION: Methods For Their Production And Use
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OSTROLENK, FABER, GRRB & SOFFEN
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,804
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/114,420
; FILING DATE: 30-AUG-1993
; APPLICATION NUMBER: US 08/112,387
; FILING DATE: 27-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,928
; FILING DATE: 08-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/612,330
; FILING DATE: 13-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gray III, William O.
; REGISTRATION NUMBER: 30,944
; REFERENCE/DOCKET NUMBER: P/1261-13
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-402-804-6

Query Match      6.9%; Score 231.5; DB 2; Length 242;
Best Local Similarity 28.9%; Pred. No. 2.5e-15;
Matches 61; Conservative 46; Mismatches 95; Indels 9; Gaps 5;

QY      80 GYKTLKLSGKFCRRRELIGMGPSGAGKSTFMNLAGYSGMKGOILVN---GRPRE 135
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      12 GTHVHLKTFNLVSKGKGLVIGPSGSGKSTTIRCMNGLEEVS-SGEVVVNLVNLHKK 70

QY      136 LRTFRKMSYINQDDMLPLHLTVLEAMVYSANLNTENPDVKNLDLVEITLTALGLMSCSH 195
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      71 IEICRYKAMVQHFENLYPHMTVLQNLTLAPMKLQKKKKEAETAFKYLKVVGLLDKAN 130

QY      196 TRTALLSGORKELATALELVNPPVPMFFDEPTSGILDSASCFQVYSLMSLA-QGRTII 254
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      131 VYPATLSGGQQQKQKVAIARSLCTKKPYILFDEPTSLDPTTQEVLDVMKEISHQSNNTMV 190

QY      255 CTHOPS-AKLFEMFDKLYLSOGOCIFKGV 284
      | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      191 VVTHEMGFAK--EVADRIIFMEDGAIVEENI 219

RESULT 12
US-08-772-270A-12
; Sequence 12, Application US/08772270A
; Patent No. 6019984
; GENERAL INFORMATION:
; APPLICANT: MacInnes, Janet
; APPLICANT: Ricciatti, Paul
; APPLICANT: Mallard, Bonnie
; APPLICANT: Rosendal, Soren
; TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR
; NUMBER OF INVENTION: PRODUCING SAME, AND THEIR USE AS VACCINES
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bereskin & Parr
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2

```

Matches 60; Conservative 37; Mismatches 81; Indels 3; Gaps 2;
QY 80 GYKTLKCLSGKFCRRRELIGTMGSPGAGKSTFMNLAGYRESGMKGQILVNGRPRELRTF 139
DB 12 GKOILSDFSUSIEKQILALVGPSSGSKTLLRLMAGL-ETIDSGQIFYNGQPLEDEL 70
QY 140 RMSC--YIMODMLLPHLTVLEAMVMSANLNTENPDVKNLDTLVTILALGLMSCSHTR 197
DB 71 QKRNLGPFVDFQLFPHLSVLENTLSPVKTMGKQEEAEKASGLLEQLGLGHAESY 130
QY 198 TALLSGGORKRLATALELVNPNPFDEPTSGDLSASCFOVWSLMKSLAOGGRTICTI 257
DB 131 PFSLSGGOKORVALARAMIDPEITIGYDEPTSLDPELRLEVERKLILQNRRELGMTQIVVT 190
QY 258 H 258
DB 191 H 191

RESULT 7
US-09-134-001C-3832
Sequence 3832, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134.001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3832
LENGTH: 242
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3832

Query Match 7.1%; Score 238.5; DB 4; Length 242;
Best Local Similarity 30.3%; Pred. No. 4.8e-16;
Matches 67; Conservative 45; Mismatches 94; Indels 15; Gaps 5;
QY 77 RRGYKTLKCLSGKFCRRRELIGTMGSPGAGKSTF---MNLAGYRESGMKGQILVNGR 133
DB 11 KKGANEVLRDINTVKEGVVAILGSPGSKSTLLRCMNL---DVPSGKGVIFEDNE 66
134 R-----ELRTRKMSCYIMQDDMLLPHLTVLEAMVMSANLNTENPDVKNLDTLVTILAL 188
67 LTQNVHLDNLRQKMGVWFQNFPHKVKVIENVVNLAPLLHLKDSKDLKEKALYILEKV 126
QY 189 GLMSCSHTRALLSGGORKRLATALELVNPNPFDEPTSGDLSASCFOVWSLMKSLAQ 248
DB 127 GLKDKADSYNQLSGGOKQORVAIALAMEPDVMLFDEPTSLDPELRLEVERKLILQNR 186
QY 249 GGRITICTIHOPS-AKLFEFMDKLYLSOGOCIFKGVVTNL 288
DB 187 ESMTMVIVTHMNFAR--EISDKVFMADGVVVSGETPQNI 225

RESULT 8
US-09-134-001C-3215
Sequence 3215, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134.001C
CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3215
LENGTH: 225
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3215

Query Match 7.1%; Score 236.5; DB 4; Length 225;
Best Local Similarity 28.9%; Pred. No. 6.9e-16;
Matches 65; Conservative 47; Mismatches 94; Indels 19; Gaps 6;

QY 61 IEFVELSYSVREGPCWKRKGYKTLKCLSGKFCRRRELIGTMGSPGAGKSTFMNLAGYRE 120
DB 7 LEFQVSKSFRDGNQMIQAVKPTDLKIQ-----ELIAIVGSPGSKSTFLT-MAGALQ 60
QY 121 SGMKGOILVNGRPRELRTFKMS-----CYIMQDDMLLPHLTVLEAMVMSANLNTEN 173
DB 61 RPTSGNIYINKNISILSEKOLQIRINEIGFILQSTNLVPLFIEQOKL---LGKYYK 117
QY 174 PDVKNLDTLVTILALGLMSCSHTRTALLSGGORKRLATALELVNPNPFDEPTSGDLS 233
DB 118 DTLSEDEVQKLVKQNLSDITNQLPNOISGGOKQORVAIAKAIYTNPSIILADEPTASLD 177
QY 234 ASCFOVVS-LMKSLAOGGRTICTIHOPS-AKLFEFMDKLYLSOG 277
DB 178 NNAMAVMKILOKQTKERNKTCVIVTH--DERLTHFCDKTFMEDG 220

RESULT 9
US-09-134-001C-3731
Sequence 3731, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134.001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3731
LENGTH: 248
TYPE: PRF
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3731

Query Match 7.0%; Score 235.5; DB 4; Length 248;
Best Local Similarity 28.5%; Pred. No. 1e-15;
Matches 70; Conservative 47; Mismatches 110; Indels 19; Gaps 5;

QY 55 KRSADVIEFVELSYSVREGPCWKRKGYKTLKCLSGKFCRRRELIGTMGSPGAGKSTFMNI 114
DB 1 KRVSIMIELSHIHKSFNE-----TEVIKIDLKLINOGEVVTLIGRSGSGKTTLLRM 51
QY 115 LAGYRESGMKGQILVNGRPRELR-----TPRKMSCYIMQDDMLLPHLTVLEAMVMSANL 168
DB 52 INAL-EITPTEGVYVNGWYNTYNTKKKQIKVRQSGMVWFQNFPHKSALENWEGILIT 110
QY 169 NLTENPDVKNLDTLVTILALGLMSCSHTRTALLSGGORKRLATALELVNPNPFDEPT 228
DB 111 VKKNKATANEAMNLAKEVLVHVQORPHALSGGQORVAIALAMNPKVMLFDEPT 170
QY 229 SGLDSASCFOVWSLMKSLAOGGRTICTIHOPS-AKLFEFMDKLYLSOGOCIFKGVVTN 287
DB 171 SALDPELVNDVLKVIKELADEGMTMIVTHMREK--EVSNOIAFTHIEGVIAEQGTPED 228

	QY	377	LKRTFTSILRDTVLTHLRMSHWVIGVLGILLYLHIGDDASKVFNNTCGLFFSMFLMF	4336
	Db	380	VSKRSFKNLGNPOASIAQIIVTVLGLVGITAYFGLKNDSTGIQRAGLVLEFLTNOCF	4339
	QY	437	AALMPTVLTPLEMAVMREHLNWKSLAKYLAKTMD-VPEQVVCPVVCISIVYWMTG	495
	Db	440	SSVASVEL-FVVEKKLIHEVYSRYRVSYFYGLKLLSDLLPMTPLSPIFTCIYVFMUG	498
	QY	496	OPAETSRFLFSALATATALVAOSLGLLIGAASNLSQVATFYGPVTAIPVLLSFGEFVSF	555
	Db	499	LKPRADAFVNMFPLMWVAYSASSMALAIAGAOSVSVATLLMTICFVPMMIFSGLLVNL	558
	QY	556	KTIPTYLOWSSYL SVRYRGEGVILTIIY-----GME-RGDLTCLCEERCPRQPQSI LR	607
	Db	559	TTIASWLSWLQYFSIPRYGTALQHNEFLQNFCPLGNATGNPNWCYTCT-GEYLVKQ	617
	QY	608	ALDVEDAKLYMDFLVLGIFFLAURLLAYLYRYVK	643
	Db	618	GIDISPWGMLKNHVALACMTIVIFTIAYKLKLFUK	653

ULT 2

US-08-919-573-2
; Sequence 2, Application US/08919573
; Patent No. 6346392
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin, Karl Russel
; TITLE OF INVENTION: NOVEL GLUTAMINE TRANSPORT
; TYPE OF INVENTION: ATP-BINDING PROTEIN
; NUMBER OF SEQUENCE: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price, & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,573
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```
Query Match      7.6%   Score 255.5; DB 4; Length 244;
Best Local Similarity 31.1%; Pred.No.9e-18;
Matches 56; Conservative 46; Mismatches 87; Indels 13; Gaps 4;
```

```

QY      132 RPRELFRKMSCVIMODDMLLPRLPLTVLEAMVMWSANLNLTENPDVKNDLVTEILTALGLM   191
Db      72 KKNDLFAMREKGWVFQQFLNFPNMNTVIENITLSPIKTKGESRVAEKRAQLLKKEVGLP    131
QY      192 SCSHRTALLSGGORKLATALALELVNNPPVMFFDEPTSGLDSDASCFOVVSLMKLSLAOGGR 251
Db      132 DKADATPQSUGGGQQRIARGLAMPEDVLDFEPTSALDPDMVEGVIAVMDQLAKSGM     191
QY      252 TIICTHIQPSAKLFEMFDKLYILSQQGCIFKG 283
Db      192 TMVIVTMEGMGAR-EVADRVIEMAGDVVVVEDG 222

RESULT 3
US-08-919-573-4
; Sequence 4, Application US/08919573
; Patent No. 6346392
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin, Karl Russel
; TITLE OF INVENTION: NOVEL GLUTAMINE TRANSPORT
; NUMBER OF INVENTION: ATP-BINDING PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/919,573
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: P50597
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-919-573-4

```

	Query Match	7.6%	Score 255.5;	DB 4; Length 244;
	Best Local Similarity	31.1%;	Pred.	No. 9e-18;
	Matches .66;	Conservative	46;	Mismatches 87; Indels 13; Gaps 4;
Qy	80	GKTLKLCKSGKCRCRELIGIMGPSGACKSTF---MNILAGYRESGMKGQILVNG-----	131	
Dd	16	KKEVLKGINLEIKRGVVIIIGPSGSKTSLRSMNLL---EATKKVFIEGVDIID	71	
Qy	132	RPRLETFRKMSCVIOMDDMLPHLTIVLEAMMYSANLNTENPDVKNDLVTEIILTALGLM	191	
Dd	72	KKNDLFAMREKMGNVFOFNLFNMVTIENTLSPIKTGESVEAEKRAQLLKELVGLP	131	
Qy	192	SCSHTRALLSQQGRKRRLATALELVNNPPVMWFEDPTSGLDSDASCFOVSUMKSUAOGR	251	
Dd	132	EADAYPQSISGGQQRRIARTARGAMPDVLFFDEPTSAIDPEMVGEVLVAODLASGM	191	

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2003, 16:43:24 ; Search time 26 Seconds
(without alignments)
731.046 Million cell updates/sec

Title: US-10-072-621-9

Perfect score: 3347

Sequence: 1 MAEKALPAGCGGLPGAVAM.....FLALRLLYLVRYVKSR 646

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Sorted: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A-COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B-COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A-COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B-COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PTUS-COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	710	21.2	655	US-09-245-808-1	Sequence 1, Appli
2	255.5	7.6	244	US-08-919-573-2	Sequence 2, Appli
3	255.5	7.6	244	US-08-919-573-4	Sequence 4, Appli
4	247.5	7.4	270	US-09-134-001C-5561	Sequence 5561, Ap
5	243.5	7.3	256	US-09-134-001C-4600	Sequence 4600, Ap
6	238.5	7.1	203	US-08-858-207A-315	Sequence 315, App
7	238.5	7.1	242	US-09-134-001C-3832	Sequence 3832, Ap
8	236.5	7.1	225	US-09-134-001C-3215	Sequence 3215, Ap
9	235.5	7.0	248	US-09-134-001C-3731	Sequence 3731, Ap
10	233.5	7.0	341	US-09-134-001C-4042	Sequence 4042, Ap
11	231.5	6.9	242	US-08-402-804-6	Sequence 6, Appli
12	230.5	6.9	711	US-08-772-270A-12	Sequence 12, Appl
13	223	6.7	215	US-09-305-984-18	Sequence 18, Appl
14	223	6.7	215	US-09-073-341A-18	Sequence 18, Appl
15	222.5	6.6	236	US-09-134-001C-4595	Sequence 4595, Ap
16	222.5	6.6	707	US-08-772-270A-4	Sequence 4, Appli
17	220	6.6	1307	US-08-395-246C-2	Sequence 2, Appli
18	218	6.5	215	US-09-305-984-64	Sequence 64, Appl
19	218	6.5	224	US-09-305-984-72	Sequence 72, Appl
20	218	6.5	224	US-09-305-984-74	Sequence 74, Appl
21	218	6.5	224	US-09-305-984-76	Sequence 76, Appl
22	215.5	6.4	254	US-09-134-001C-4440	Sequence 4440, Ap
23	215	6.4	1334	US-08-996-545-2	Sequence 2, Appli
24	215	6.4	1334	US-09-328-320-2	Sequence 2, Appli
25	214.5	6.4	246	US-09-627-376-15	Sequence 15, Appl
26	213.5	6.4	382	US-09-134-001C-3439	Sequence 3439, Ap
27	212.5	6.3	244	US-09-134-001C-3641	Sequence 3641, Ap

28	206.5	6.2	202	4	US-08-858-207A-332	Sequence 332, App
29	206	6.2	1684	3	US-08-665-259-25	Sequence 25, Appl
30	206	6.2	1684	3	US-08-762-500-25	Sequence 25, Appl
31	206	6.2	1704	3	US-08-762-500-75	Sequence 75, Appl
32	205	6.1	475	4	US-09-212-247C-4	Sequence 4, Appli
33	204.5	6.1	233	4	US-09-627-376-12	Sequence 12, Appl
34	204.5	6.1	1302	1	US-08-232-537-2	Sequence 2, Appli
35	203.5	6.1	791	1	US-08-394-880B-2	Sequence 2, Appli
36	202.5	6.1	1349	2	US-08-612-734B-2	Sequence 2, Appli
37	201.5	6.0	1279	2	US-08-784-649A-2	Sequence 2, Appli
38	201.5	6.0	1280	2	US-08-583-276-19	Sequence 19, Appli
39	201.5	6.0	1280	6	5206352-4	Patent No. 5206352
40	200.5	6.0	337	4	US-09-134-001C-5550	Sequence 5550, Ap
41	199	5.9	257	4	US-09-134-001C-3539	Sequence 3539, Ap
42	198.5	5.9	1280	2	US-08-752-447-2	Sequence 2, Appli
43	198.5	5.9	1280	4	US-09-316-167-2	Sequence 2, Appli
44	195	5.8	460	4	US-09-134-001C-3369	Sequence 3369, Ap
45	193	5.8	376	2	US-08-997-080-89	Sequence 89, Appl

ALIGNMENTS

RESULT 1

US-09-245-808-1
; Sequence 1, Application US/09245808
; Patent No. 6313277
; GENERAL INFORMATION:
; APPLICANT: Doyle, L. Austin
; APPLICANT: Abruzzo, Lynne V.
; APPLICANT: Ross, Douglas D.
; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which
; FILE REFERENCE: encodes it
; FILE REFERENCE: Ross UMB conversion
; CURRENT APPLICATION NUMBER: US/09/245.808
; EARLIER APPLICATION NUMBER: 60/073763
; EARLIER FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Human MCF-7/AdrVp cells
US-09-245-808-1

Query Match	21.2%	Score 710;	DB 4;	Length 655;
Best Local Similarity	27.8%	Pred. No. 2e+63;		
Matches 177;	Conservative 140;	Mismatches 247;	Indels 72;	Gaps 16;
QY	61	IEFVELSYSV--REG--PCWRKRGYKTLKCLSGKFCRRELIGIMGPSGAGKSPFMNLA	116	
Db	37	LSFNINICRYKLSGGLPC-RKPEKEILSNING-IMKPGNLALGPTGGKSLLDVLA	94	
QY	117	GYRE-SGMKQGLVNGRPRELRTFRKMSYIMQDDMLPHLTVLEAMMVSAANLML--TEN	173	
Db	95	ARKDPSGLGDLVINGAPRP-ANFKCNSGVYQDDVVMGTLTVRENLFQSAALRLATMT	153	
QY	174	PDVKNDLVTELLTALGLMCSHRTAL-----LSGGORKRLATALELVNPPVFFDEPT	228	
Db	154	NHEKNERINRVIQELGLDKVADSKVGTQFIRGVSGGERKRTSIGMELITDPSILFDEPT	213	
QY	229	SGLDASCFQVSVLMKSLAOGRTICTIHOPSAKLFEMFDKLVILSGOCIFKGVVTNL	288	
Db	214	TGLDSSANAVLLLLKRMKSGQRTILFISIHOPRYSIFKLFDSLTLLASGRMLFHGPAQEA	273	
QY	289	IPYLKGLGLCHPTYHNPADFIIEVASGEYGLNPLMFRAVQNGLCAMAEKKSPKNEVP	348	
Db	274	LGVFESAGYHCEAYNPNADFDLIINDG-----STAVALNREEDFKATEII	319	
QY	349	APCPPCPEVDPI-----ESH-----TATSTLTQFCI	376	
Db	320	EPSQDKPLIEKLAELIYVNSSFYKTRKAEHLHLSGGEKKKKTIVFKETISYTTTSFCHLRW	379	